Università degli Studi di Ferrara

DOCTORATE OF PHYLOSOPHY IN
BIOMEDICAL SCIENCES AND BIOTECHNOLOGY

CYCLE XXXI

COORDINATOR Prof. Pinton Paolo

MOUSE TDP-43 EXPRESSION DURING DEVELOPMENT IS EPIGENETICALLY MODULATED

SSD BIO/10

PhD student
Pacetti Miriam

Tutor
Prof. Pagani Franco

Years 2015-2018

SUMMARY

LIST OF ABBREVIATIONS

Aß	Amyloid-ß
AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
AML	Acute myeloid leukemia
CFTR	Cystic fibrosis transmembrane conductance regulator
CHMP2B	Charged multivescicular body protein 2B
CK	Casein kinase
CMML	Chronic myelomonocytic leukemia
CpG	CG di-nucleotides
CTF	C-terminal fragments
DM	Myotonic dystrophies
DNMT	DNA methyl transferase
DPR	Dipeptide repeat
EAAT2	Excitatory amino acid transporter 2
eIF3	Eukaryotic initiation factor 3
ER	Endoplasmic reticulum
ESE	Exonic splicing enhancer
fALS	Familiar amyotrophic lateral sclerosis
sALS	Sporadic amyotrophic lateral sclerosis
FTLD	Frontotemporal lobar degeneration
FUS	Fused in sarcoma
G3BP	RasGAP-associated endoribonuclease
GPNMB	Glycoprotein nmb
HAT	Histone acetyltransferases
HD	Huntington's diseas
HDAC	Histone deacetylases
HIV1	Human immune-deficiency virus type 1
HMT	Histone methyltransferases
hnRNA	Heterogeneous RNA
hnRNP	Heterogeneous ribonucleoprotein
IBD	Inclusion body myopathy

LCS	Low complexity sequences
MBD	Methyl-CpG-binding domain
MCT1	Monocarboxylate transporter 1
MDS	Myelodisplastic syndrome
MeCP2	Methyl-CpG binding protein 2
mRNP	messenger ribonucleoprotein particles
ncRNA	Non coding RNA
NEFH	Encoding neurofilament heavy polypeptide
NGS	Next generation sequencing
NES	Nuclear export signal
NLS	Nuclear localization signal
NMJ	Neuromuscolar junction
NTF	N-terminal fragment
NXF1	Nuclear RNA export factor 1
PABP	PolyA-binding protein
P-bodies	Processing bodies
POMA	Paraneoplastic opsoclonus myoclonus ataxia
Q/N	Glutamine/Asparagine
PUF	Pumilio/FBF
QKI	QUAKING
RAN	Repeat-associated non-AUG
RBC	Red blood cell
RBP	RNA binding protein
RBD	RNA binding domain
RGG	Arginine-Glycine-Glycine
RNP	Ribonucleoprotein particle
RRM	RNA recognition motif
RT	Retro Transcription
SAM	S-adenosyl-L-methionine
SG	Stress granule
sHsp	Small heat shock protein
SOD-1	Superoxide dismutase 1
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particle
STX1B	Syntaxin 1B

SUMO	Small ubiquitin-like modifications
SR	Serine-Arginine
TAF15	TATA-binding protein associated factor 15
TDG	Thyamine-DNA glycosylase
TDPBR	TDP-43 binding region
TDP-43	TAR DNA-binding protein 43
TIA-1	T-cell intracellular antigen 1
TET	Ten-eleven translocation
TLS	Traslocated in liposarcoma
tRNP	Transport ribonucleoprotein particles
TSS	Transcription start site
PrP	Prion protein
VCP	Valosin-containing protein

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease with a median age of onset of 50-60 years. 95% of ALS cases show inclusions of an aggregated form of the hnRNP TAR DNA-binding protein 43 (TDP-43), resulting in its nuclear clearance and loss of function. Previously our laboratory has developed an aggregation model based on the TDP-43 "prion like" sequence and it was shown that the aggregates induced are capable of sequestering endogenous TDP-43 (or TBPH in the case of *Drosophila*) and inducing loss of function. Particularly interesting for our studies was the observation that there was a late onset of a locomotion defects in the fly that coincided with an age-related reduction of TBPH protein and mRNA levels in the brain. This suggests that the pathology onset may, in part, be determined by the physiological age related decrease in expression of TDP-43. In fact it can be hypothesized that the continuous sequestering of newly synthesized TBPH/TDP-43 by the aggregates becomes critical only when the amount of newly synthesised TDP-43 can no longer compensate this process. This physiological drop has been observed to be evolutionary conserved in *Drosophila* and mice. We have decided to extend the analysis of TDP-43 expression during development to different ages and tissues in mouse and other organisms. We found that TDP-43 showed different trends of expression that are time- and tissue- specific. In fact, even if ubiquitously highly expressed at the early post natal developmental stage in brain and some peripheral organs like liver and skeletal muscle, expression decays dramatically at 90 days after birth in the skeletal muscle, mildly in brain and does not vary in liver. We also observed this peculiar age related changes in other splicing related RNA binding proteins. The phenomena deserve further analysis and we have focused in the mechanism responsible for the lower transcription rate. Epigenetic studies in mice revealed that DNA methylation levels of the *TARDBP* promoter have tissue- and developmental stage- specificity that have an inverse correlation with TDP-43 expression. In particular, in 90 days aged mice, the 4th CpG island of *TARDBP* promoter displays an increase in the methylation rate of the majority of its CG sites in the skeletal muscle where TDP-43 dramatically decrease while a lower number of CG sites increased their methylation rate in brain where the protein mildly decrease during development. On the other hand no significant methylation rate changes have been detected in liver where TDP-43 is sustained during time. We also found that the highly methylated CG sites of the cultured mouse motorneurons NSC34 *TARDBP* promoter perfectly match with the CG sites significantly methylated in mice. Furthermore, 5-azacytidine demethylating treatment of this cell line leads to an increase in the level of

TDP-43 mRNA and protein expression in a dose- and time- dependent manner. Accordingly, through an *in vitro* methylation assay we observed that mouse *TARDBP* promoter significantly decreased its activity in NSC34 cells. Moreover, we found that point mutagenesis to prevent methylation of the specific CG dinucleotides that we found as the most methylated in the promoter fourth island leads to an increased *TARDBP* mouse promoter activity. Interestingly, 5-azacytidine demethylation treatment of the human neuronal cell line SH-SY5Y has the same effect of increasing TDP-43 mRNA production observed in mouse motorneurons.

In conclusion, our data suggest that mouse TDP-43 mRNA and protein levels are regulated in a tissue specific and time dependent manners and this behaviour can eventually be extended to other members of the RNA binding proteins family. Changes of *TARDBP* promoter methylation status in the different mice tissues during development and aging are inversely correlated with TDP-43 expression and preliminary data suggest that *TARDBP* promoter methylation could be a pivotal player also in the regulation of human TDP-43.

INTRODUCTION

1. Aging and neurodegeneration

According to recent estimates of the United Nations report on world population aging, there is a global increase in life expectancy that will lead the number of people aged 60 and older worldwide to more than double in the next 35 years, reaching almost 2.1 billion[1]. The improvement of the standard and expectancy of life will also result in an increase in neurodegenerative disorders. Indeed, almost all aged brains share characteristics with neurodegeneration and this has brought into discussion the question of whether these hallmarks could represent the precursors for neurodegenerative disorders[2].

Aging, which is fundamental to neurodegeneration and dementia, is defined as a dynamic, multisystemic and highly heterogeneous process that progressively leads to decline of fitness over time, and finally to death[3]. It involves multilevel changes resulting in organ dysfunction, tissue deterioration and disorganization, and loss of stem cell renewal. From the viewpoint of the molecular biology, aging mirrors gradual deterioration of molecular components and the related events regulating cell viability and proliferation[4]. Although aging arises with different phenotypes among tissues, there are some common hallmarks such as the genomic and epigenetic instability, mitochondrial dysfunction, telomere attrition and the accumulation of senescent cells[5] which are all associated with deep changes deregulating several processes such as cytokine secretion, alteration of gene expression and alternative splicing as well as chromatin remodelling. Senescence is considered one of the causes leading to tissue deterioration during aging and it is an irreversible process of programmed cell-cycle arrest happening in most diploid cell types[6]. The progressive accumulation of senescent cells occurs in aging tissues and it has been hypothesised whether this can lead to the enhancing of the senescence of neighbouring cells by a mechanism called paracrine senescence[7]. While the link between senescence and aging is still unknown, it has become increasingly clear that cellular senescence plays a role in some age-related diseases and in tissue degeneration associated with aging. Consistently, the reduction of the senescent cells rates is associated with the decrease in the incidence of age-related disorder[8].

As changes in the global cellular and molecular shape are a feature of aging, it is fair to expect the breakdown in the correct gene expression and regulation. In fact, both DNA methylation and RNA metabolism are largely affected by aging. Changes that occur with aging are likely to impact the activity of factors that control splicing because of the

complexity of the regulatory machinery controlling splicing decisions. Furthermore, several studies have shown that among the age-associated deregulated processes[9] there is a prevalence of those pathways involving RNA processing and mRNA binding factors, and it is estimated that 50% of all age-associated alterations in alternative splicing are due to changes in splicing factors expression.

Since neurons are non-dividing cells with a particularly complex RNA metabolism, thus more vulnerable to the alterations occurring during aging, there is a strong overlap between aging and neurodegeneration. However, there are important differences in the severity of the alterations in the pathologic and not-pathologic condition. What is not known is why the aging brain progresses to neurodegeneration in some cases but not in others. In fact, several autopsy studies have confirmed the presence of amyloid plaques, neurofibrillary tangles, Lewy bodies, inclusions of TAR DNA-binding protein 43 (TDP-43), synaptic dystrophy, the loss of neurons and of brain volume in brain of people who had not been diagnosed with a neurological disease. Whether these lesions are precursor of the neurodegeneration process rather than a physiological event occurring during aging is still unknown[2]. The hypothesis that defects in proteostasis contribute to or are a result of aging or of neurodegeneration is becoming increasingly popular even if it still not fully elucidated. During the normal aging also the macromolecules oxidation exceeds the physiological levels leading to their accumulation and consequential digestion failure by the lysosomes that are no longer able to eliminate the cellular material leading to overproduction of lysosomial enzymes[10]. In general, both normal aged and diseased brains present increased levels of lysosomal proteins and enzymes, and neurons are characterised by impaired and abnormal endosomes, lysosomes and autophagosomes.

However, one limit in distinguishing the significant proteins abnormalities in the brain is given by the discrepancy between clinical manifestation of dementia and its associated physical characteristics in the brain. For instance, studies on cerebrospinal biomarkers for amyloid-β and tau show that around 30% of people cognitively classified unimpaired are positive for these predictive markers of disease meaning they could be at a preclinical and asymptomatic stage of the pathologic condition of neurodegeneration[11]. In conclusion, the diagnosis correlates with relevant pathological features only in few cases. The majority of old individuals suffer for those processes that eventually lead to neurodegeneration, however some of them might have compensatory mechanisms to overcome the produced cellular defects and to maintain normal cognition.

It is likely that neurodegeneration and disease are the consequences of the normal brain aging and that the typology of the neurodegenerative disorder is determined by the

individual genetic and environmental background of factors occurring during time[12] (**Figure 1**). In order to understand how age contributes to disease, several studies have been performed by intersecting the key genetic and environmental factors of the disease with particular hallmarks of ageing and by identifying the importance of that process in the disease[13].

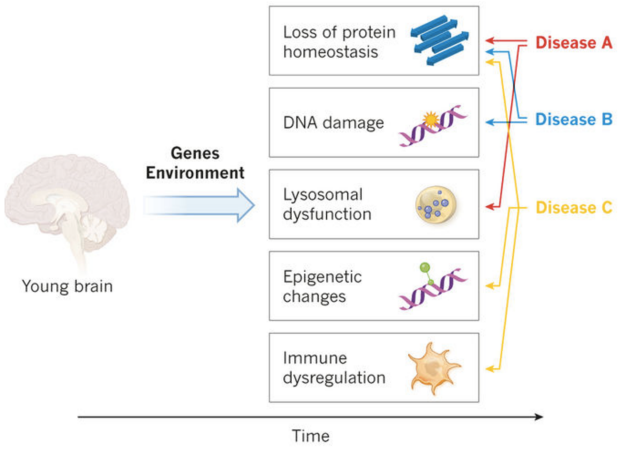

Figure 1. Molecular pathways, genetics and environmental contribution to aging and neurodegeneration over time. Aging can be dissected into individual processes, including a loss of protein homeostasis, DNA damage, lysosomal dysfunction, epigenetic changes and immune deregulation. The individual genetic background and the environmental factors exposure, determine the kind and the susceptibility to develop a certain disease over time in accordance to the spatiotemporal distribution of the lesions. From Tony Wyss-Coray, 2016

Twin studies have assessed that the factors that more influence human lifespan are acquired with age, and the heritability plays only a minor role[14]. On the other hand, the environmental factors including lifestyle, diet, exposure to toxins, drugs of abuse represent the 70% of variation in lifespan[15].

Another aid in the understanding the contribution of aging to neurodegeneration is given by the study of gene expression, DNA methylation and other epigenetic DNA modifications that change dramatically with age[16]. For example, it has been found that healthy people show a specific age-related brain molecular signature in term of changes in gene expression that is identified in almost all the individual with Frontotemporal lobar degeneration (FTLD) or Alzheimer's disease (AD), independently from their age[17]. Furthermore, methylation analysis in AD autopsied brains in two independent studies showed a correlation of the pathology with an increased DNA methylation of the major genes linked to the disease[18,19]. Genetic and epigenetics studies can therefore help to uncover the molecular pathways that link aging with neurodegeneration. Broadly speaking, omics studies on brain diseased regions will probably help to gain better insights into new targets to delay disease-prone changes of aging.

2. Neurodegenerative disorders

Neurodegenerative diseases are characterized by progressive dysfunction and loss of neurons leading to a wide range of symptoms including motor dysfunction, cognitive failure and dementia[20]. It is increasingly recognized that many of these diseases share the fundamental phenomenon of the presence of proteins with altered physiochemical properties deposits in human brain[21]. Indeed, in these disorders a core concept is the fundamental role of the unfolded proteins[22] that alter the response of the protein elimination pathways, such as the ubiquitin-proteasome system and the autophagy-lysosome pathway, stress response proteins and chaperones[23]. These pathways interact with other pathways leading to dysregulation of energy metabolism, molecular damage, metabolic changes, dysregulation in ion homeostasis, and adaptation[24]. All these alterations result in the accumulation of one or more neuronal proteins triggering cell death and neuronal loss[25]. Even though these diseases are characterized by the accumulation of insoluble aggregates of normally soluble proteins in the central nervous system, there are striking phenotypic differences among them. This is due to the different distribution of deposition of pathologic aggregates in the brain, the plethora of morphological distribution and protein composition of the deposits and the various cellular hosts or extracellular localization of the aggregates[26]. Overall, on the basis of these concepts, every neurodegenerative disorder might be characterized by taking into consideration the following aspects: (1) the major clinical symptoms determined by the anatomical region showing neuronal dysfunction or loss; (2) the typology of the proteins that show conformational change and biochemical modifications and (3) the cellular and subcellular pathology meaning which kind of neural cell type (also including the cellular compartment) shows pathological protein deposits or whether these are found extracellularly[27] (**Table 1**). Accordingly, to each neurodegenerative syndrome an anatomical, cellular and protein vulnerability can be assigned, that acts together with the individual genetic background to influence the susceptibility to develop these diseases (**Figure 2**)[28].

Despite the proposed mechanism of disease progression, the exact causes that trigger the formation of these inclusions are still unclear. Given that for each neurodegenerative disease the major component of misfolded protein is different, or in some cases includes combinations of several proteins, it is thought that aggregates are an end-stage manifestation of the disease, ending with the collapse of protein homeostasis-dependent cellular maintenance[29].

Disease	Location	Aggregated protein
Alzheimer's	Extracellular	Amyloid-B
	Intracytoplasmic	Tau
	Intracytoplasmic	a-synuclein
Amyotrophic lateral sclerosis	Intracytoplasmic	Superoxide dismutase-1 (SOD-1)
	Intracytoplasmic	TAR-Binding Protein 43 (TDP-43)
	Intracytoplasmic	Sarcoma/Translocated in lipodsrcoma (FUS/TLS)
Frontotemporal Dementia	Intracytoplasmic	Tau
	Intracytoplasmic	TAR-Binding Protein 43 (TDP-43)
	Intracytoplasmic	Sarcoma/Translocated in lipodsrcoma (FUS/TLS)
Cortical basal degeneration	Intracytoplasmic	Tau
Dementia with Lewy bodies	Intracytoplasmic	a-synuclein
Huntington disease	Intranuclear	Hungtintin
Multiple system atrophy	Intracytoplasmic	a-synuclein
Parkinson's disease	Intracytoplasmic	a-synuclein
Pick's disease	Intracytoplasmic	Tau
Prion diseases	Extracellular	Protease-resistant prion protein (PrP)
Spinocerebellar ataxia	Intracytoplasmic	Ataxin

Table 1. Neurodegenerative diseases characterization on the basis of major protein costituent of the aggregates and their cellular localization.

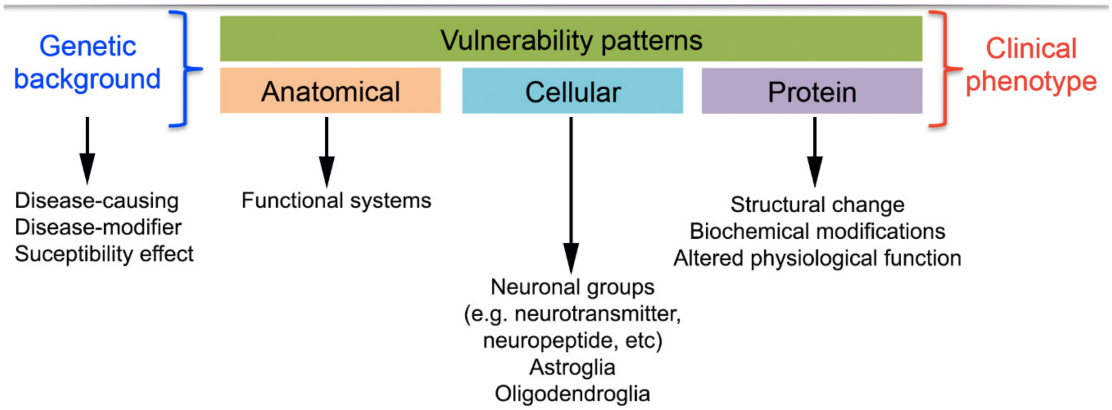

Figure 2. Vulnerability patterns in neurodegenerative diseases. Each coloured box represents different vulnerability pattern. Together they act determining the individual clinical phenotype. From Gabor G. Kovacs, 2016

Considering this ideology, a myriad of studies have been performed to identify the pathological mechanism behind neurodegenerative disorders. While some analyses were focused on the identification of mutations of the misfolded proteins, others, using the new techniques of next-generation sequencing (NGS), have attempted to establish a pattern of polymorphisms able to confer vulnerability in developing the disease[30]. Research from these studies have found that the proteins more significantly associated with neurodegenerative disorders are[31]: (1) The microtubule-associated protein Tau, encoded by the *MAPT* gene; (2) Amyloid-ß (Aß), which is cleaved from a large transmembrane precursor protein (Aß-precursor protein or AßPP); (3) α-synuclein, which is encoded by the *SNCA* gene; (4) Prion protein (PrP), encoded by the gene *PRNP* on the chromosome 20; (5) Transactive response (TAR) DNA-binding protein 43 (TDP-43), an highly conserved nuclear protein, encoded by the *TARDBP* gene on chromosome 1[32]; (6) Fused in sarcoma protein (FUS), Ewing's sarcoma RNA-binding protein1 (EWSR1), and TATA-binding protein associated factor 15 (TAF15), collectively called as FET proteins[33]; (7) the proteins encoded by genes linked to neurological trinucleotide repeats disorders (e.g., huntingtin, ataxins, atrophin-1), neuroserpinopathy, ferritin-related neurodegenerative diseases and familial cerebral amyloidosis. Based on the major components of the aggregates in the central nervous system, neurodegenerative proteinopathies are classified as tauopathies, a-synucleinopathies, TDP-43 proteinopathies, FUS/FET proteinopathies, prion diseases, trinucleotide repeats diseases, neuroserpinopathiy, ferrinopathy, and cerebral amyloidoses (**Table 2**).

However, for the majority of cases which are sporadic or idiopathic, mutations and polymorphisms in relevant genes cannot account for the aberrant folding of the proteins or the increased propensity for aggregation[34,35] as they are in the wild type form. Indeed, most

neurodegenerative diseases are classified as complex diseases that seem to be linked to both genetic risk and environmental factors, ranging from decreased efficiency in cellular processes attributed to aging, to harmful chemical exposures[36]. Thus, whilst on the one hand several attempts have been done to gain better insights into the genetics of the neurodegeneration, on the other hand, different research areas have taken a more molecular approach trying to identify the involved cellular pathways and the related targets. This is the case, for example, of the pathological mechanism of the Huntington's diseas (HD) and the myotonic dystrophies (DM) in which the tri-nucleotide repeat expansions lead to RNA toxicity[37,38]. These discoveries identified the RNA as a toxic species, thus expanding the neurodegenerative research field into the role of RNA metabolism and RNA binding proteins (RBPs) in neurodegenerative diseases[38].

Disease Group	Protein	Disease Type	Form	Phenotype
AD	Tau, Aβ	AD	SP/GEN	DEM
Tauopathy (FTLD-Tau *)	Tau	PiD	SP	FTD
		GGT	SP	FTD
		CBD	SP	MD/FTD
		PSP	SP	MD/FTD
		AGD	SP	DEM
		NFT-dementia/PART	SP	DEM
		FTDP-17T	GEN	FTD/MD
TDP-43 proteinopathy	TDP-43	FTLD-TDP (type A–D)	SP/GEN	FTD
		MND-TDP	SP/GEN	MD
		FTLD-MND-TDP	SP/GEN	FTD-MD
FUS (FET)-proteinopathy FTLD/MND-FUS (FET)	FUS/FET	aFTLD-U, NIFID, BIBD	SP	FTD/MD
		MND-FUS	GEN	MD
α-Synucleinopathy	α-Synuclein	PD	SP/GEN	MD
		DLB	SP/GEN	DEM/MD
		MSA	SP	MD
Prion disease	PrP	sCJD, VPSPr, sFI	SP	DEM/MD
		iCJD	ACQ	DEM/MD
		vCJD	ACQ	DEM/MD
		Kuru	ACQ	DEM/MD
		gCJD, GSS, FFI, PrP-CAA	GEN	DEM/MD
TRD **	Huntingtin	HD	GEN	MD
	Ataxin 1, 2, 3, 7, CACNA1A, TBP	SCA 1, 2, 3, 6, 7, 17	GEN	MD
	FMRP	FXTAS	GEN	MD
	ARP	SBMA	GEN	MD
	Atrophin-1	DRPLA	GEN	MD
Other forms	Ferritin	Hereditary ferritinopathy	GEN	DEM/MD
	Neuroserpin	Neuroserpinopathy	GEN	DEM
	ABri, ADan, gelsolin, cystatin, transthyretin, Aβ	Hereditary amyloidoses/CAA	GEN	DEM
	Only UPS	FTLD-UPS	GEN	FTD
	Not determined	FTLD-ni	SP	FTD
	Tau, α-Synuclein	NBIA	GEN	DEM/MD
	Tau, a-Synuclein, TDP-43	Various genetic and sporadic diseases ("secondary" proteinopathy forms)	SP/GEN	DEM/MD

Table 2. Summary of most relevant modifications of neurodegeneration-related proteins with remarks. From Gabor G. Kovacs, 2016

Abbreviations: ACQ: acquired; ARP: androgen receptor protein; AD: Alzheimer disease; AGD: Argyrophilic grain disease; BIBD: Basophilic inclusion body disease; CAA: sporadic cerebral amyloid angiopathy; CACNA1A: α1A subunit of the P/Q-type voltage-gated calcium channel (SCA6; Cytoplasmic aggregates); CBD: Corticobasal

degeneration; CJD: Creutzfeldt-Jakob disease (i: iatrogenic, s: sporadic, v: variant, g: genetic); DEM: dementia; DLB: Dementia with Lewy bodies; DRPLA: dentatorubral-pallidoluysian atrophy; sFI: sporadic fatal insomnia; FFI: fatal familial insomnia; FMRP: Fragile X mental retardation protein; FTD: frontotemporal dementia; FTLD: frontotemporal lobar degeneration; aFTLD-U: atypical FTLD with ubiquitinated inclusions; FTLD-UPS: FTLD with inclusions immunoreactive only for the components of the ubiquitine proteasome system; FTLD-ni: FTLD no inclusion specified; FTDP-17T: Frontotemporal dementia and parkinsonism linked to chromosome 17 caused by mutations in the *MAPT* (tau) gene; FXTAS: Fragile X associated tremor and ataxia syndrome (here also astroglial inclusions); GEN: genetic; GGT: globular glial tauopathy; GSS: Gerstmann-Sträussler-Scheinker disease; HD: Huntington disease; INIBD: intranuclear inclusion body diseases; MD: movement disorder; MND: Motor neuron disease; MSA: multiple system atrophy; NBIA: neurodegeneration with brain iron accumulation; NIFID: Neurofilament intermediate filament inclusion disease; PD: Parkinson disease; PiD: Pick disease; PrP: prion protein; PSP: Progressive supranuclear palsy; SCA: spinocerebellar ataxia; SBMA: spinal and bulbar muscular atrophy; SP: sporadic; TBP: TATA-Box binding protein (SCA17); TRD: trinucleotide repeat expansion disorder: refers to genetic disorder and associated with different proteins; VPSPr: variably protease-sensitive prionopathy. * FTLD is not typical in AGD or PART; ** only SCAs with protein inclusions.

3. RNA metabolism in neurodegenerative disease

Diverse disease mechanisms are responsible for neurodegeneration. Although more interest has been placed on the role of protein aggregates, numerous approaches have also highlighted the fact that altered RNA processing is a contributing factor in the pathogenesis of these diseases[39–42]. Aberrations can occur at all levels of gene regulation, RNA synthesis, function and degradation. These can result in stoichiometric differences of the RBPs that subsequently can lead to their mis-localization and sometimes sequestration into aggregates. Furthermore alterations can also occur at microRNA biogenesis level as well as changes of long-intergenic non-coding RNAs[43].

Aging contributes to a significant extent in determining the profound change of molecular pathways involved in gene regulation. As an example, several studies have identified age-related alterations in splicing factors expression[17]. In particular, there is an age-related deviation from normal splicing patterns due to altered expression of factors responsible for the alternative splicing outcome[44]. Further investigation will likely provide new insights into the gene expression changes that distinguish healthy aging from neurodegenerative profile.

3.1 Neurodegeneration and RNA-binding proteins

3.1.1 Major classes of RNA binding proteins and significance in gene expression regulation

There are many levels of regulation of gene expression occurring from the generation of the primary RNA transcript from DNA to translation into protein. Post transcriptional gene regulation depends on RNA binding proteins (RBPs) acting on several aspects of mRNAs life such as splicing, mRNA cellular localization, stability, turnover, polyadenylation,

nuclear export, editing and the fine tuning of association with the translational machinery[45,46] (**Figure 3**). These functions are carried out through the interactions between RNA and proteins participating in the ribonucleoprotein complex formation.

RBPs contain one or, more often, multiple RNA-binding domains (RBDs) through which they interact with their targets[47]. Furthermore, auxiliary functional domains contribute in achieving a higher degree of modularity which creates both RNA-binding and functional diversity within the RBPs [48]. Finally, alternative splicing and post-translational modifications such as phosphorylation, arginine methylation and small ubiquitin-like modifications (SUMO)[46], generate additional layers of complexity, as they can modify the RNA-binding, function and localization of the ribonucleo protein (RNP) [49]. Additionally, growth factors, oxidative stress and other stimuli can alter the phosphorylation status. It should also be taken into account that many macromolecular complexes contain specific combinations of RBPs that may be cell- or tissue type-specific; furthermore many RBPs are expressed with cell-, tissue- and/or developmental-type specific patterns. All these factors contribute to increase the variability of their functional profiles[45].

It is estimated that around 2-8% of the total number of genes are likely to encode for a RBP in eukaryotes[50,51]. This could be an underestimation as the bioinformatics predictions are based on the binding motifs and very likely there are additional undiscovered RBDs. In any case, this is a considerable number of genes, which gives rise to an even higher number of protein variants following post-translational, splicing, alternative splicing, and combinational modifications occurring during generation of each RBP.

Families of RBPs are defined on the base of sharing common features even though they differ in domain composition and functional properties. Among them, heterogeneous nuclear ribonucleoproteins (hnRNPs) are a well characterized class of nucleoplasmic-localized multifunctional RBPs that can regulate RNA splicing, export, localization, translation, and stability[52].

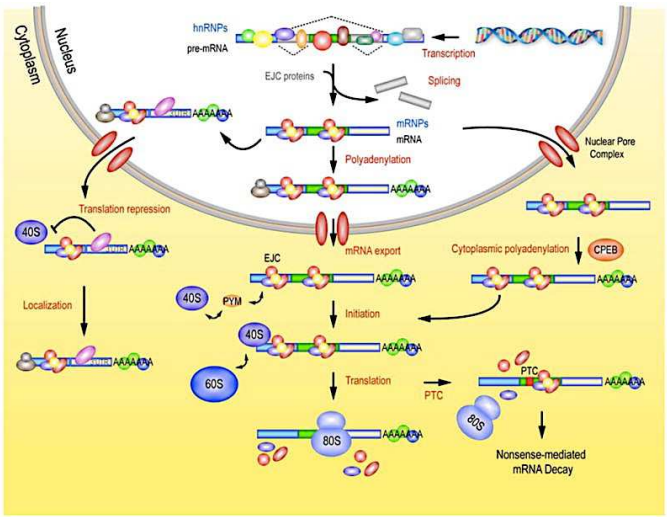

Figure 3. Functions of RBPs at transcriptional level. After transcription in the nucleus, RNA is subjected to different processing steps that can determine the fate of the transcript. Splicing events affects several following steps, such as recruitment of ribosomal subunits for translation initiation, as well as surveillance of mRNA for nonsense-mediated mRNA decay (NMD). Some mRNAs might be additionally modified by the cytoplasmic polyadenylation RNP, CPEB. From Glisovic T, et al., 2008

Because of the size, heterogeneity, and localization of the eukaryotic nascent transcripts, when refereeing to pre-mRNA the term heterogeneous nuclear ribonucleic acids (hnRNAs) is utilised, and since these hnRNAs are associated with proteins forming complexes before their maturation, collectively, these hnRNA-binding proteins are named hnRNPs[53]. Four unique RBDs were identified in hnRNPs: the RNA recognition motif (RRM) which is the most diffused, the quasi-RRM, a glycine rich domain constituting an RGG box, and a KH domain. Next to RBDs, hnRNPs frequently contain auxiliary domains, such as proline-, or glycine- rich domains[54]. Over 50% of hnRNPs have been characterized to play a role in splicing[55]. Broadly speaking, usually hnRNPs exert a splicing suppression effect [56,57], however, there are cases in which the opposite tendency takes place[53]. Regarding their role in cytoplasmic RNA metabolism, hnRNPs are involved in transport of mRNA, translation, and stability [58,59,60,61].

Another family of RBPs is structurally defined by the modular domain structure consisting of one to two amino- terminal RNA recognition motifs (RRMs) and a carboxyl-terminal domain rich in serine (S) and arginine (R) dipeptide repeats, thus named Serine and Arginine-rich (SR) proteins[62]. There are twelve canonical members of the SR protein family that share this domain structure[63]. Dynamics of the SR protein localization in the cell is largely influenced by RS domain phosphorylation, acting also in the regulation of their activity. Mechanistically, this kind of post-translational modification allows the SR protein to gain a major RNA-binding specificity and it is also important for specific protein-protein interactions within the pre-spliceosome. Structurally, phosphorylation results in the decrease of the disordered RS thus promoting the formation of more ordered side chain for molecular recognition[64]. In the nucleus, they function as molecular adaptors linking the pre-mRNA to the splicing machinery by associating with nascent RNA transcripts, thereby facilitating the spliceosome assembly. In alternative splicing, SR proteins generally work as enhancers binding the exonic splicing enhancers (ESEs) elements [65] thus counteracting the effect of several member of the hnRNPs that act mostly as splicing silencers. Furthermore, it has also been found an interaction between SR and the canonical mRNA export factor nuclear RNA export factor 1 (NXF1) suggesting a role of these RBPs in the nucleocytoplasmic shuttling of mRNAs[66]. However their depletion does not result in general defects in mRNA nuclear export, thus leading to the hypothesis

of a redundancy effect of other factors. In addition to splicing, nuclear export and translation, SR proteins also influence mRNA stability[67].

Spliceosome components also belong to the RBPs family. They are termed as snRNPs (small nuclear RNPs) because of their association to a set of small nuclear RNAs (snRNAs) acting together to remove the non-coding regions (introns) from the pre-mRNA transcripts[68].

Gene expression regulation is a critical process involved in a wide range of cellular functions in both pathological and physiological conditions. This event dysregulation is a feature of a multitude of neurological age-related diseases[69]. In the following paragraphs it will be outlined the relevance of the splicing regulator proteins in a much more specific neurodegenerative context.

3.1.2 Role of neural-specific RNA binding proteins in neurological diseases

More than 1000 genes in eukaryotic genome encode for RBPs and, among them, more than 50% are neuronal-specific supporting the concept that neurons have developed a particular system for the RNA metabolism modulation[70].

Genetic studies have revealed mutations in or dysregulation of RBPs in a wide range of human neurological diseases. For instance, the Hu/ELAV protein family expression outside the nervous system leads to paraneoplastic neurological syndrome[71]. The disease is caused by the production of autoantibodies against the Hu family proteins triggered by ectopic expression of RBPs physiologically expressed only in the nervous system[72].

Another example of how the ectopic expression of these neural-specific proteins leads the organism to recognize them as “non self” is the one of Nova1/Nova2 proteins. In breast cancer, their aberrant ectopic expression causes autoantibodies production and, as a result the autoantibodies affect the regions of the central nervous system where Nova proteins are normally expressed causing paraneoplastic opsoclonus myoclonus ataxia (POMA)[73].

The importance of RBPs for neural function is further highlighted by pathological alterations generated by their mutations. In fact, several RBPs genetic alterations have been associated to neurological diseases, among them those of the Rbfox family. Rbfox 1 is expressed in neurons, heart and muscle. In brain tissues Rbfox 1 is expressed only in neurons[74] and mutations in the Rbfox1 gene are associated with neurons-affected phenotype such as mental retardation, epilepsy and autism[75]. Furthermore, CNS-specific conditional Rbfox 1 knock out mice exhibited spontaneous seizures[76].

Pathological reduced expression of an RBPs can also result in disease. One example is the regulation of differentiation of myelin-forming oligodentrocytes and Schwann cells by the QUAKING (QKI) protein[77]. Expression of QKI mRNA is reduced in disease-related regions, and multiple studies suggest it as a key player in several psychiatric diseases including schizophrenia and ataxia[78].

In other cases, the pathogenic mechanism that leads to disease is a gain of toxicity of the transcript because of a repeat expansion mutation. This is the case of the abnormally expanded CGG triplet repeats in the 5' UTR of the *FRM1* gene, encoding an RNA-binding protein associated with X-Fragile syndrome[79], causing hypermethylation of the gene promoter, leading to transcriptional silencing[80].

3.1.3 RNA-binding proteins in neurodegeneration

All nascent RNA molecules need to be correctly processed, addressed, stabilized and degraded. All these tasks are carried out by the RBPs that are the key players of both coding and non-coding eukaryotic RNAs. They perform their roles through extensive protein-protein and protein-RNA interactions in several spatial and temporal combinations, that allow the control of gene expression in response to a range of stimuli[81]. This is particularly important in cells with a complex RNA metabolism, such as neurons. Furthermore, as neurons are non dividing cells, thus particularly vulnerable to any kind of deregulation, and given the intricacy of functions that the central nervous system must regulate, it is not surprising that disruption of precise RBP stoichiometry through mutations and/or pathological events can lead to a wide range of neurological and neurodegenerative diseases[45]. Evidence of this comes from recent findings showing misregulation and mutations in genes involved in RNA metabolism or the autophagy/proteasome pathways, which play an important role in the onset and progression of several neurodegenerative disorders. Moreover, there is a high level of regulation in the management of the transcripts transport along neurons. As noted before, the RRMs and glycine-rich domains allow the RBPs to create a network of interactions between other proteins and transcripts making RNA handling, transport and local translation possible[82]. This is achieved by forming organized ribonucleoprotein particles (mRNPs) such as tRNPs, processing bodies and stress granules which exert different roles[83] listed in the next paragraphs. Any kind of perturbation of these networks could potentially lead to a degenerative process.

tRNP granules

Modifications of synapses and the long lasting plastic changes in nervous system need local expression of proteins which require translocation and local transport of specific mRNA through the intricate dendritic arbours before starting to be translated[82]. To address this need of motile and protective transport, specific ribonucleoprotein particles (tRNP granules) are constantly formed[84]. These complexes contain silent mRNAs, RBPs (including FMRP, hnRNP A2, FUS/TSL, TDP-43, TAF15, staufen1, staufen2), densely packed clusters of ribosomes and miRNAs[85,86]. Once the tRNP reaches the proper compartment, the translation can start.

The possible role of tRNA granules in neurodegeneration is not well established. However, mutations and/or alterations of the component proteins may result in a kind of stress for the neuronal management of translation. Furthermore, it has been described that there are ALS-associated mutations in TDP-43 that induce changes in the protein transport direction within axons[87].

Processing bodies

mRNAs can also be organized in mRNP granules which serve as pools of mRNAs. These multi-molecular assemblies are called mRNA-silencing foci and they contain repressed mRNAs, miRNAs and RBPs working as RNA silencers. The two main mRNA-silencing foci are the processing bodies (P-bodies) and the stress granules (SGs).

P-bodies (PBs) are a dynamic structure that contains mRNA decay machinery components. In fact, because of their enrichment in translation repressors, components of the mRNA decay machinery, RNA decapping machinery, and RNA-mediated silencing components, they are thought to play a role in the mRNA storage and degradation[88]. PBs are highly dynamic multi-molecular assemblies with factors that recruit mRNA through multimerization domains. They are often observed juxtaposed to SGs, with which they exchange mRNAs, even though they are also present in un-stressed cells. PBs have been also suggested to interact with tRNPs[86].

Stress granules: physiology and connections to aging and neurodegeneration

SGs represent one way adopted by the cell to achieve the energy preservation and to protect macromolecules in stressful conditions, such as hypoxia, heat-shock, chemical exposure, and aging[89]. In adverse conditions, the cell facilitates its own survival by

prioritizing the synthesis of enzymes and chaperones required for stress adaptation and shutting down the translation of housekeeping genes. At the same time these non-membranous structures are formed in a reversible manner in order to sequester the released mRNAs and RBPs and keep them silent and protected from degradation until the stress subsides[90].

SGs are highly dynamic structures and in normal conditions, their quick disassembling results in the rapid translation restoration. Furthermore, the fact that these granules are directly connected with P-bodies and in dynamic equilibrium with polysomes facilitates the shuttling of RNA and proteins from one spatially localized compartment to another[84].

At molecular level, SGs are composed by several RBPs such as the T-cell intracellular antigen 1 (TIA-1), RasGAP-associated endoribonuclease (G3BP), eukaryotic initiation factor 3 (eIF3), and polyA-binding protein (PABP)[91].

Although initially small, with prolonged stress time, the SGs start to increase in size eventually leading to aggregation foci formation. In fact, the RBPs accumulated inside are dynamically associated with mRNAs and are thought to promote their aggregation.

Increasing genetic and experimental evidences support the interpretation sustaining SGs dynamic participates in pathogenesis of age-related neurodegenerative diseases[92]. First, one of the signatures of these disorders is the accumulation of cytosolic protein/RNA aggregates and these cytosolic inclusions are enriched in nuclear RNA-binding proteins normally recruited in SGs upon stress[93,94]. Second, several disease-linked mutations which increase the protein propensity to aggregate have been identified in genes encoding SGs components[95,96]. Third, the pathogenic expansion of a hexanucleotide repeat in the *C9orf72* gene involved in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTLD), induces the production of toxic dipeptide repeats (DPR) proteins via repeat-associated non-AUG (RAN) translation. These DPRs interact with low complexity sequences (LCSs) in RBPs, thus altering the dynamics of membrane-less organelles, such as nucleoli, and SGs[97].

Moreover, RNPs contain stable and long-lived structures that can be disassembled only by autophagy. Given that, this clearance pathway is physiologically impaired and slowed down with aging and it is likely that in young individuals there is a better response to stress stimuli which therefore result to be much more tolerated than in older people[98]. However, this is not the only mechanism proposed to trigger the aggregation process in neurodegeneration, indeed several other pathways are thought to play a fundamental role in the pathogenesis of these disorders.

3.2 Changes in splicing factors expression and alternative isoforms are associated with aging and neurodegeneration

As mentioned above, aging leads to a generalized impairment in cellular plasticity and adaptability. For example, the autophagy process in elderly people is not able to fully face the need of clearance after stress condition, resulting in induced aggregation of un-cleared proteins. Among all the theories of aging, there is one considering the accumulation of damage to DNA over a lifetime as associated with large scale alterations of gene expression but to date, there are little data to assess this[99]. However, considering the alterations observed on the RNA metabolism in old population studies, it is likely this theory has fundamentals. One of the most evident alterations is the one regarding alternative splicing in general, which arises from the age-related modification of splicing factors expression.

A variety of age-related expression analyses in cell lines or stored cell materials have been reported, but their reproducibility is limited because of the small sample sizes and of the sensitivity of mRNA transcripts to variation of storage and handling[100]. Moreover, the majority of cases only consider blood leukocytes, human brain and senescent fibroblasts, but little is known about the other tissues. These studies shown an age related deregulation of splicing patterns maybe attributable to changes in the pool of regulatory splicing factors and, to a lesser extent, of the core splicing factors[44]. The specific identity of the affected factors and their level of modification is not always the same in different tissue types, suggesting different degrees of tissue specificity. Specific splicing factors expression has also been associated with longevity in rodents and humans[3]. The consequences of deregulation of splicing factors expression with age remains to be determined, but it is likely that there is a gradually loss of ability in regulating alternative splicing that results in the imbalance of mRNA isoforms abundances and in the inability to properly react to stress stimuli.

Changes in splicing factors and alternative splicing are particularly evident in age-related disease. Alteration of splicing has been reported in relation to AD and Parkinson's disease. In particular, there is an age-related gene expression signature as a marker of the increased genetic susceptibility of developing dementia[101,102]. Brains from AD patients showed changes in splicing factors expression: a reduction of PTBP2 and RAVER1 and an increase in ROD1 and PTBP1, and the pattern of changes observed in the pathology fits with the expression signature of aging. Specific disease associated mutations in critical

splicing genes and aberrant mRNA processing are also documented in AD[17]. Furthermore, Tauopathies are also a family of neurodegenerative diseases where altered alternative splicing contributes in determining the accumulation of filamentous Tau isoforms[103,104]. Similarly, in ALS and a large number of FTLD cases deregulation of splicing events[105–107] and aberrant aggregation of the splicing factor TDP-43 in the cytosol of affected neurons is observed [108,109].

A lot of work is still required to identify the pathological mechanisms leading to the deterioration of the physiological pattern and abundance of splicing factors and how this can impact on aging and age related diseases. However, all the studies described above provide a link between neurodegeneration, aging and the processes of alternative splicing, which modulate the adaptability, plasticity, and cell identity. Aberrations in alternative splicing might be the cause of the high degree of heterogeneity of elderly population and this might be a starting point to better understand age related diseases, many of which highly depend on correct splicing process[110].

4. The epigenetics of aging and neurodegeneration

Epigenetics is a quickly growing field encompassing mechanisms regulating gene expression without changes in genomic sequence. It is increasingly assuming a central role in neurological disorders. In fact, epigenetic deregulation plays a pivotal role in aging that is one of the major risk factor for developing neurodegeneration. Importantly, variations in epigenetic regulation could provide a reasonable explanation to the profound changes in the gene expression landscape characterizing neurodegenerative diseases[111].

Epigenetic regulation covers multiple levels of gene expression: it can regulate transcription rate via acting directly on DNA and histone tails as well as translation rate via creation of interactions with mRNAs[112,113]. Importantly, aging and age-related diseases include defined changes in 5-methylcytosine content and are generally characterized by genome-wide hypomethylation and promoter-specific hypermethylation[16]. Because of their reversibility, epigenetic modifications represent one of the preferential targets for the therapeutic approaches. Indeed, several drugs specifically targeting DNA methyltransferases (DNMTs) are subjects of clinical trials and DNMTs, together with associated partners, showed great promises for clinical treatments[16].

4.1 Basic mechanisms

Histone modifications

The nucleosome is the basic unit of chromatin and consists in DNA wrapped around an octamer of histones (H2A, H2B, H3, and H4)[114]. The histone tails protrude from the nucleosome core and are subjected to different post-translational modifications including serine and threonine phosphorylation, lysine acetylation, lysine and arginine methylation, lysine sumoylation and ubiquitination, and proline isomerization[115]. Histone modifications thus became a target for a variety of chromatin associating/remodelling factors, enhancing or restricting the access of regulatory proteins to DNA[116].

The histone acetyltransferases (HAT) and histone deacetylases (HDAC) promote the addition or removal of the acetyl group in the lysine residues of the histones. Acetylated lysine are binding platforms for protein containing the bromo-domain, also present in some transcription factors, as well as in subunits of chromatin-remodelling complexes[117]. Overall, the acetylation of lysine residues of the histones leads to transcriptional activation[118]. By contrast, the removal of the acetyl group from the histone tails is associated to chromatin condensation and heterochromatin assembly, blocking the transcriptional machinery from acceding to the DNA resulting in transcription silencing.

Additionally, histone methyltransferases (HMTs) catalyse the transfer of the methyl group to lysine or arginine residues of histone H3 and H4. The effect of the histone methylation depends on the residues that are modified, it could result in repression (methylation of H3K9, H3K27, and H4K20) or activation (methylation of H3K4, H3K36, and H3K79) of transcription[119].

Non-coding RNAs

Non-coding RNAs (ncRNAs) are the most recently identified players in the epigenetics field. The best characterized ncRNAs are miRNAs. They are small ncRNAs acting at post-transcriptional level, silencing their targets. Even if they do not alter the chromatin structure, they are considered as mediators of the epigenetic mechanisms of regulation since they allow changes in gene expression without alterations in the DNA sequence[120].

Long ncRNAs are an additional class of non-coding RNAs which play an important role in the regulation of transcription via their interaction with chromatin or chromatin-associated factors[121,122].

DNA methylation

DNA methylation is an evolutionarily conserved system of transcriptional repression present in both prokaryotic and eukaryotic organisms[123,124]. The methylation reaction is catalysed by DNA methyltransferases (DNMTs), enzymes that transfer a methyl group from S-adenosyl-L-methionine (SAM) to the C5 of a cytosine residue, creating 5-methylcytosine (5-mC)[16] (**Figure 4**). Methylation concerns almost exclusively the cytosines within CG di-nucleotides (CpGs), which are approximately methylated in the 60%-90% of cases in mammalian genomes with the exception of CpGs within active gene promoter regions that are normally un-methylated. Even if recently other non-CpG methylation forms have been described, little is known about them. Apart from cytosine, also guanine and adenine, even if more rarely, are susceptible to methylation, resulting in 7-methylguanine (7-mG) and 3-methyladenine (3-mA), respectively[125]. In this thesis, however, DNA methylation refers exclusively to 5-mC, unless mentioned otherwise.

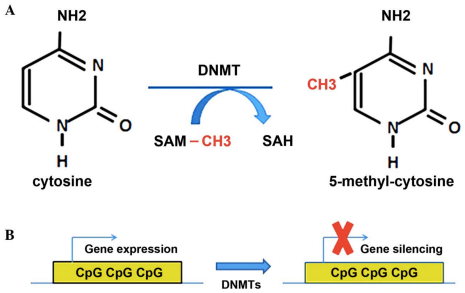

Figure 4. DNA methylation reaction. (A) DNMT catalyzes the trasfer of the methyl group from a molecule of SAM to the C5 of a cytosine. (B) Methylation is generally associated to gene silencing. From Xiangrong C., et al. 2016

The methyl group addition to cytosines within CG di-nucleotides prevents transcription factors or the transcriptional machinery to recognise their signature sequences. In fact, the majority of the mammalian transcription factors has DNA recognition elements containing CpG-rich motifs as well as CG-rich binding sites. Thus, DNA methylation acts as a steric obstruction that prevents the binding of these factors. Moreover, the addition of a methyl group can also alter nucleosome stability or position directly excluding the access of the transcriptional machinery[126].

Additionally, methyl-CpG-binding domain (MBD) proteins are involved in gene silencing via interacting with methylated promoters. These proteins show increased binding affinity for methylated DNA over un-methylated and through their binding to methylated DNA,

they recruit co-repressor complexes able to establish a permanent condition of chromatin silencing due to their interaction with histones deacetylases[127]. Five major mammalian MBDs proteins exist: methyl-CpG binding protein 2 (MeCP2), MBD1, MBD2, MBD3, and MBD4[128].

Although still a point of discussion, some of the mechanisms proposed for active DNA methylation/demethylation include base excision repair, enzymatic removal of the methyl group, and oxidative demethylation of 5mC[129]. The most compelling scheme of DNA demethylation supposes it is initiated by the oxidation of 5-mC into 5-hydroxymethylcytosine (5-hmC) by ten-eleven translocation (TET) proteins. TET enzymes can further oxidize 5-hmC to 5-formylcytosine, then to 5-carboxylcytosine (5-caC), and this last can be changed back to cytosine by decarboxilation[130].

Alternatively, 5-caC nucleotides can be directly excised by thyamine-DNA glycosylase (TDG). In supporting this, ablation of TDG results in abnormal DNA methylation patterns, and on the other hand, overexpression of the enzyme induces a decrease in the genomic content of 5-caC. Furthermore, ESC lysates lose glycosylase activity against 5-caC in the absence of TDG[131,132].

Alternatively, hydroxymethylation by TET may also passively demethylate DNA by disrupting the ability of DNMTs and MBD proteins to access the previously nonhydroxilated 5-mC[133]c.

4.2 DNA methylation changes during aging

DNA methylation plays a crucial role during development. Indeed, the establishment of an epigenetic clock, based on a correlation observed between age and methylation, allows a prediction of age from the analysis of methylation profiles. More precisely, age-dependent DNA methylation shows a logarithmic relationship with chronological age until adulthood, and a linear relationship later in life, indicating the epigenetic clock “ticks” faster during growth and development[111].

In parallel, twin studies have shown the existence of an epigenetic drift caused by environmental factors or spontaneous stochastic errors in the process of transmission of DNA methylation that leads to unpredictable differences in the methylome among aging individuals[134].

Overall, mammalian aging is associated with CpG hypomethylation in loci outside CpG islands, and this is likely to contribute to the loss of heterochromatin over time. On the contrary, an age-dependent hypermethylation usually occurs at specific CpG sites of the

genome, presumably to repress expression of specific genes (**Figure 5**)[135–137]. However, even if particularly accurate in most tissues, a significantly lower methylation status can be distinguished in tissues with a high turnover of renewal[111].

With the advent of novel next-generation sequencing technologies, several studies have been performed to assess the genome-wide assessment of DNA methylation, confirming and extending previous studies. If generally the age-related methylation changes have been documented in a more prominent way in CpG islands, tissue specific biases are frequently observed[138]. Furthermore, these changes occur in a highly reproducible manner even if their relevance has not yet been fully explored and understood.

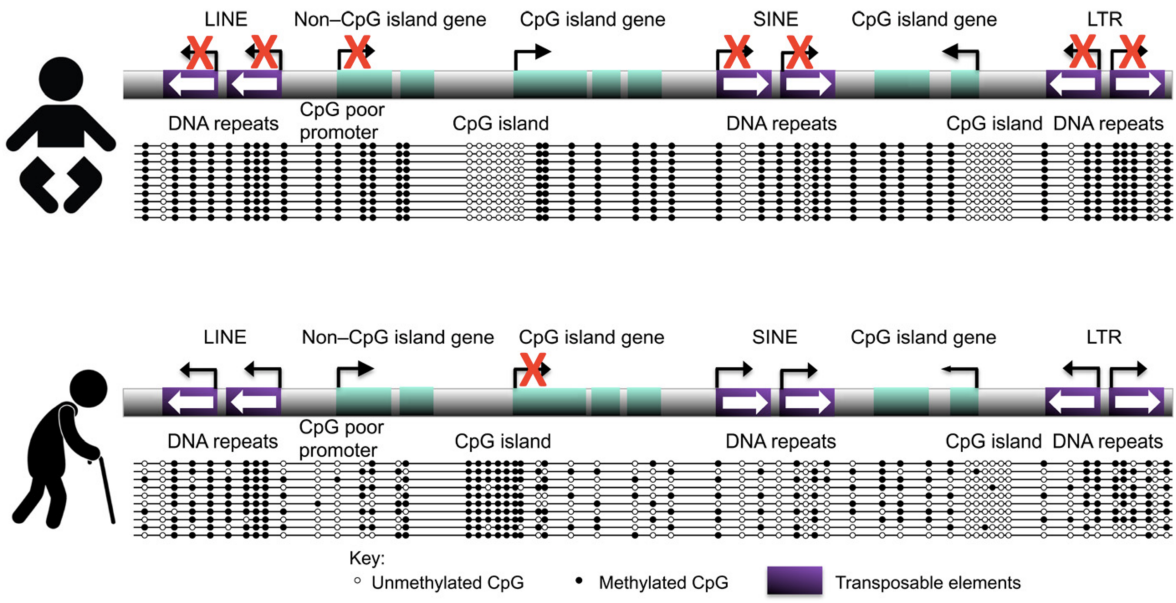

Figure 5. Schematic representation of DNA methylation changes during aging. Young mammalian cells show a generalized DNA hypermethylation over the genome, with the exception of CpG islands within promoter of actively transcribed genes. The heavy methylation of DNA repeats (LINE, SINE, LTR) helps the maintainance of the consitutive heterochromatin state at that level. During aging, a general DNA hypomethylation occurs in a stochastic manner, leading to the activation of DNA regions normally silenced like transposable elements. However, DNA methylation specifically increases in CpG islands of genes leading to their silencing. From Sangita P., at al. 2016

The further discovery that there is an altered pattern in the DNA methylation status that allows the accurate prediction of the cell passage number provides a link between replicative senescence and aging[139].

A multitude of *in vivo* studies have also divulged the relevance of DNA methylation during aging. A study performed via pyrosequencing probed 217 free-disease human tissues from 10 anatomic sites, and dissected DNA methylation of 1413 autosomal CpG loci associated with 773 genes in both young and old individuals[140]. The authors found CpG island-dependent connections between methylation and age. In particular, loci outside the islands showed decreased methylation rate, whereas loci within the islands gained methylation rate with age, as previously observed[140]. By the analysis of the peripheral blood DNA in a

population study including middle and advanced aged people, Bellizzi et al. found hypomethylation to be associated with frailty and loss of function. A 7-years follow up showed a worsening in the health status related to a 5-mC decreased content[141]. Several studies have also outlined the link between cancer and methylation. Notably, local hypermethylation events during aging frequently occur at the promoter of tumor-suppressor genes[142]. Moreover, studies on normal human prostate tissues from 45 donors revealed hypermethylation in genes such as *RARb2, RASSF1a, GSTP1, NKX2-5,* and *ESR1*, which harbour mutations believed to be a risk factor for cancer[143].

Other studies have focused the attention on how DNMTs modulation might play a central role in longevity. In fact, flies with overexpressed *dDnmt2*, the only methyltransferase *Drosophila* gene, show a boost of up to 58% in mean life span compared to controls. In the same flies, multiple small heat shock protein-encoding (sHsp) genes (*Hsp22, Hsp23*, and *Hsp26*) were also upregulated[144]. Conversely, sHsp-encoding genes were downregulated in flies with diminished levels of *dDnmt2*[144]. This study demonstrates that, in *Drosophila*, *dDnmt2* is a regulator of life span and stress resistance.

In summary, there is a generalized trend of global hypomethylation and CpG islands specific-hypermethylation occurring over time. This phenomenon has also been detected in various organs in rats[145] and mice[146] where hypomethylation was found to be more pronounced. The decrease in 5-mC content could lead to less efficient gene regulation, and the CpG hypermethylation to an inappropriate silencing of specific genes. Such changes may induce genome instability, thus contributing to aging phenotypes.

4.3 DNA methylation and age-related neurodegeneration

While in the past data implicating DNA methylation came more frequently from cancer studies, numerous age-related pathologies have now been linked to aberrant 5-mC content. For example, elevated amyloid-β peptide (Aβ) levels have been linked with an AD-specific hypomethylation of the *APP* gene promoter region[147]. AD patients are also characterized by decreased cerebrospinal levels of folate and SAM and, concomitant with this, an overall hypomethylation status in the temporal neocortex neurons nuclei[148,149]. Remarkably, the finding that tau gene expression is also affected by hypermethylation could be relevant not only for AD but also for other age-related tauopathies. In fact, with age, the transcriptional activator SP1 binding site in the tau gene promoter region became hypermethylated inducing an age-related tau decreased expression, and this was confirmed also in human frontal cortex and hippocampus autopsies[150]. Another case of promoter-specific

demethylation has been found in patients with PD, which present some spherical masses of aggregated α-synuclein, the Lewy bodies, as a major disease mark. The disease is caused by duplication or triplication mutations in the *SCNA* gene resulting in the accumulation of the protein. Hypomethylation of the CpG island 2 within its promoter has been associated to an increased α-synuclein accumulation, and subsequently a greater aggregation and a more severe phenotype[151]. Moreover, other genes in PD patients including *PARK16*, glycoprotein nmb *(GPNMB*) and syntaxin 1B (*STX1B*) have also been reported to be differentially methylated in PD[152].

Mutations in human *DNMT1* have been recently reported to cause a form of neurodegeneration and sensory neuropathy accompanied to dementia. The disease causing mutations induce a global hypomethylation, as well as site specific hypermethylation and reduced methyltransferase activity[153].

All together, the studies over the past decade have shown changes in the epigenetic landscape linked to aging and age-related disorders. Moreover, these modifications could represent disease biomarkers and it is important to underscore the fact that they are reversible, therefore representing a tempting target for therapeutic intervention.

4.4 Epigenetic based therapies

DNA methylation, with its presumed involvement in a plethora of diseases ranging from cancer to neurodegenerative disorders, represents a promising target for therapeutics. Several approaches can be considered, including the use of nucleoside analogues, non-nucleoside analogues, nutritional modifiers (methyl-donor and bioactive components), antisense oligonucleotides, and siRNAs.

Nucleoside analogues are either modified ribonucleosides or deoxyribonucleosides incorporated into DNA instead of cytosine. They are recognized by the DNA methyltransferases as natural substrates, however when the enzymes initiate the methylation reaction, this result in the establishment of a covalent bond with DNA, thus their DNA methyltransferase function is blocked. Among the best known there is the zebularine, 5-Azacytidine (5-azaC), and decitibine[154]. Recently, azacitidine has become a valid option for the first-line treatment of patients with myelodisplastic syndrome (MDS)/ Acute myeloid leukemia (AML)[155]. The drug administration gives rise to significant prolonged survival compared with conventional care in AML and it is associated with lower risk of disease progression and higher rates of complete remission, partial remission, haematological improvement and red blood cell (RBC) transfusion independence[156].

The non-nucleoside analogues have been shown to inhibit the DNMTs activity without the need of being incorporated into DNA but, unlike the previously described molecules, the mechanism of action is not well understood.

The use of methyl donors represents a different way of modifying DNA methylation status. The replenishment with methionine and choline, as well as the coenzyme required in one-carbon metabolism (such as B-vitamin like folate) alters the concentration of SAM and consequentially it can modulate the 5-mC content[157]. Indeed, a diet supplemented with methyl-donor can tackle the prolonged exposure to γ-radiation stably maintaining the 5-mC amount in mouse[158].

Other studies are currently considering options leading to methylome-modification, such as antisense oligodeoxynucleotides and siRNA targeting DNMTs. The latter have been proved to reduce DNMTs activity with a decrease of CpGs methylation in cells[159]. However, no clinical data currently exist attesting the potential benefits of siRNA-mediated knockdown in methylome therapeutics.

In conclusion, epigenetic-based therapeutic approaches show clinical efficacy in treating certain human diseases such as cancer. However, the mechanisms of epigenetic pathological modifications occurring in several disorders are poorly understood and they need to be better investigated and clarified.

5. Linking RNA dysfunction and neurodegeneration in amyotrophic lateral sclerosis

In the last decade, altered RNA metabolism is emerging as a pivotal player in neurodegenerative diseases. A major revolution in the field of ALS came with the identification of two RBPs, TDP-43[109,160] and FUS/TSL[161,162], as major components of pathological inclusions found in ALS and FTLD brains of affected patients[109,163]. In particular, TDP-43 is known to be the key component of the pathological aggregates found in 97% of ALS patients[163].

5.1 Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is a heterogeneous neurodegenerative disease that has an estimated incidence of 1-2 individuals per 100,000 each year in most countries; the prevalence is about 5 cases per 100,000 people[164]. The disease was first documented in

1869 by the French neurologist Jean-Martin Charcot. It is also known as Lou Gehrig's disease in the United States and as motor neuron disease in the United Kingdom. It is characterized by the progressive degeneration of both upper and lower motor neurons, which leads to muscle weakness and eventual paralysis[165] due to the fact that corticospinal (upper) neurons make direct or indirect connections with spinal (lower) motor neurons, which innervate skeletal muscle and trigger their contraction[92] (**Figure 6**). Disease onset commonly occurs in mild-adulthood between 55 and 65 years old[166], however ALS might begin as early as in the first or second decade of life or could even emerge in later life[167]. The early symptoms include subtle cramping or weakness in the limbs, especially involving arms, legs, speech and swallowing or breathing or bulbar muscles, which progress to the paralysis of almost all skeletal muscles in the final stage[92]. There are subsets of motor neurons, such as those innervating the extra ocular muscles or sphincters, which are spared until late in the disease course. In the majority of cases, ALS leads to death in 3-5 years after diagnosis, although some forms of the disease have shown protracted survival[168].

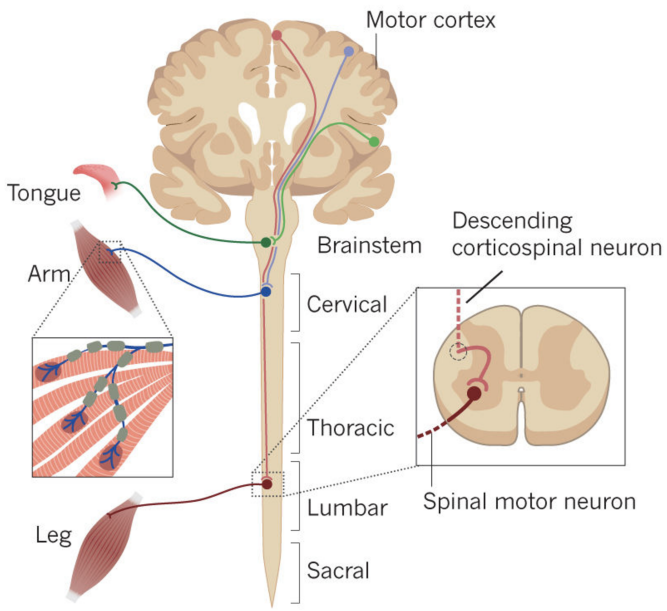

Figure 6. Nervous system components affected by ALS. Upper motor neurons project from the motor cortex into synapses in the brainstem and spinal cord, sending signals to lower motor neurons that project into skeletal muscles. In ALS patients the neuromuscolar junction is degenerated as a consequence of upper and lower motor neuronal dysfunction. Adapted from J. Paul Taylor et al., 2016

5.1.1. ALS aetiology

10% of ALS cases are familiar (fALS), and among them the trait is always transmitted with dominant inheritance and frequently with high penetrance. The first genetic mutation found to cause ALS was observed in the *SOD1* gene[169] just over 20 years ago. Since then and particularly thanks to the technological advances of NGS, several other genes carrying mutations have been identified to be implicated in ALS (**Table 3**). It is interesting to note that all these genes can be grouped on the basis of the cellular pathways they belong and on their function as follows: genes that alter proteostasis and protein quality control, genes that perturb aspects of RNA stability, function and metabolism, and genes that disturb cytoskeletal dynamics in the motor neuron axon and distal terminal[92]. The so far described mutations are mostly missense substitutions[170] even though it is important to say that the genetic alteration in *C9orf72* is something different as it is an expansion of an intronic hexanucleotide repeat[171]. Mutations in 9 of these genes (*SOD1, TARDBP, FUS/TLS, C9ORF72, OPTN, VCP, UBQLN2, SQSTM1* and *PFN1*) account for the etiology of two-thirds (68%) of fALS cases with SOD1 accounting for 1-3% and *C9orf72* expansions for 5%[171].

Locus	Gene	Protein	Protein Function	Mutations	Proportion of ALS	
					familial	sporadic
21q22.1	*SOD1*	Cu/Zn superoxide dismutase	Superoxide dismutase	>150	20%	2%
2p13	*DCTN1*	Dynactin	Component of dynein motor complex	10	1%	<1%
14q11	*ANG*	Angiogenin	Ribonuclease	>10	<1%	<1%
q36	*TARDBP*	TDP-43	RNA-binding protein	>40	5%	<1%
16p11.2	*FUS*	FUS/TLS	RNA-binding protein	>40	5%	<1%
9p13.3	*VCP*	Valosin-containing protein	Ubiquitin segregase	5	1–2%	<1%
10p15-p14	*OPTN*	Optineurin	Autophagy adaptor	1	4%	<1%
9p21-22	*C9ORF72*	C9ORF72	Possible guanine nucleotide exchange factor	Intronic GGGGCC repeat	25%	10%
Xp11.23-Xp13.1	*UBQLN2*	Ubiquilin 2	Autophagy adaptor	5	<1%	<1%
5q35	*SQSTM1*	p62	Autophagy adaptor	10	<1%	
17p13.2	*PFN1*	Profilin-1	Actin-binding protein	5	<1%	<1%
12q13.1	*HNRNPA1*	hnRNPA1	RNA-binding protein	3	<1%	<1%
5q31.2	*MATR3*	Matrin 3	RNA-binding protein	4	<1%	<1%
2q36.1	*TUBA4A*	α-tubulin 4a	Microtubule subunit	7	<1%	<1%
22q11.23	*CHCHD10*	Coiled-coil-helix-coiled-coil-helix domain containing 10	Mitochondrial protein of unknown function	2	<1%	<1%
12q14.1	*TBK1*	TANK-binding kinase 1	Regulates autophagy and inflammation	10	?	?

Table 3. The genetic of ALS. From J. Paul Taylor et al., 2016

There is also a strong interest in finding genetic variants of susceptibility to ALS. For example, it has been shown that the expansion of CAG repeats in *ATXN2* increase the risk of developing ALS[172]. By contrast, there are variants in the axonal guidance gene *EPHA4* able to determine the reduction of its expression improving the overall survival of people with ALS[173].

Considering the overall genetic environment involved in ALS etiology, even in the case of the known Mendelian-inherited genes, familial form are often characterized by <50% penetrance associated to genetic pleiotropy, with evidence of polygenic inheritance in individuals with seemingly sporadic disease[174,175].

Attempts have also been undertaken in determining environmental factors acting together with lifestyle that might be identified as causes and/or risk factors conferring higher susceptibility to develop ALS. Early epidemiological studies suggest a possible role of neurotoxins and related cyanotoxins and exposure to cyanobacterial bloom contaminated water as contributors to the increased risk of ALS in susceptible individuals[176,177]. Furthermore, the evidence of increased frequency of ALS cases among groups of athletes than in general population[178]. In addition, preliminary data from one on-going analysis on large case-control study suggest that smoking exposure might increase the risk of developing ALS and, by contrast, exposure to female contraceptive hormones and high levels of circulating lipids seem to have protective properties [179].

5.1.2 Possible pathogenic mechanisms of ALS

Despite the progress in recapitulating ALS in animal models, several limitations raised especially because usually based on gene overexpression. Nevertheless, these investigations underline that the cellular disruption characterizing the disease is the result of several different mechanisms (**Figure 7**) that synergistically culminate in larger network collapse and cannot be summarily simplified in a model[180]. It is unclear how each factor contributes to the disease, nonetheless, each pathologic mechanism is worthy to be considered as an input for future therapeutic initiatives[165].

Impaired protein homeostasis

Collapse of proteostasis has been implicated in the aetiology of a number of neurodegenerative diseases, including ALS (**Figure 7a**). In fact, mutations in disease-associated genes lead to the production of misfolded proteins with abnormal cellular localization. These can directly or indirectly impair the proteasomal or autophagic

machinery of the cell, with the consequential deregulation of the normal protein turnover[181]. Indeed, a large part of ALS-associated genes encode for proteins involved in the ubiquitin-proteasome and autophagic systems. For example, C9orf72 is a key regulator of autophagy initiation[182] and mutations in *SQSTM1* might disrupt the delivery of autophagic substrate to the autophagosome[183]. Furthermore, the fact that both SOD1 and TDP-43 are known substrates of autophagy might explain the pathologic accumulation of these proteins in ALS once the autophagic system is impaired[184]. Additionally, mutations in these genes have also been associated in deregulation of chaperone proteins and it has been shown that mutant SOD1 reduces the expression of components of the ubiquitin-proteasome system[185].

Defects in nucleocytoplasmic trafficking

Converging evidences suggest that abnormal nucleocytoplasmic trafficking is also affected by the repeat expansion in *C9orf72*. The GGGGCC repetition leads to an increased binding of mRNA export adaptors thus making the expanded pre-mRNA a target for the nuclear export, translation and production of abnormal dipeptide repeat proteins species[186]. Depletion of nuclear export adaptor reduces the production of dipeptide repeat proteins and neurotoxicity[187]. Moreover, defects in the nucleocytoplasmic transport of RNA and proteins were found in neurons derived induced pluripotent stem (iPS) cells of people with C9 ALS[188].

Axonal disorganization and disrupted transport

The peculiar architecture of the neurons allows the response to neuronal input through the alteration of local gene expression and changes in the synaptic environment that is mediated by the transport of all the necessary components for translation to distal sites for local protein synthesis[189]. Because of the asymmetric nature of these cells, axonal transport is fundamental to deliver all the components in each compartment[190]. Disorganization of the axonal cytoskeleton, and especially of the neurofilaments, is a feature of both fALS and sALS cases[92] (**Figure 7e**). Indeed, many of ALS causative mutations have been identified in proteins that are essential for axonal transport, like DCTN, PFN1, and TUBA4A[191–193]. In addition, mutations in *NEFH* (encoding neurofilament heavy polypeptide) have been described in some patients[194]. Mutations were also found in *PRPH*, encoding a cytoskeletal neural protein involved in protein cargo traffiking[195]. It has also

been demonstrated that ALS-causing mutations in TDP-43 impair the axonal transport of RNA granules in *Drosophila* models and in cultured neurons[87].

Glia cells degeneration

Studies on mutant SOD1 underlined the role of glial cells in the degeneration and death of motor neurons.

Microglia, which are the innate immune cells of the nervous system, become activate in all ALS types[196]. Mutant SOD1 drives microglia to produce high levels of extracellular superoxide, thus reverting their normal cellular function of clearance of this oxygen reactive species (**Figure 7b**). In fact, once SOD1 is misfolded, it is able to bind the GTPase RAC1 which constitutively activates the NADPH complex that produces superoxide[197]. Aberrant microglia function has also been observed as a result of mutations in *C9orf72*. This gene encodes a potential guanine exchange factor whose inactivation results in microglia dysfunction and consequential chronic neuroinflammation[198].

Another cell type showing a crucial role in the SOD1 pathogenesis is the oligodendrocytes, which are cells that myelinate the axons of upper motor neurons and the axonal initial segment of lower motor neuron. In normal conditions, oligodendrocytes are replaced by their abundant precursors. They provide the vital metabolic support to the axons through the transport of lactate via the monocarboxylate transporter 1 (MCT1) (**Figure 7c**). Mutants of SOD1 impair the expression of this transporter, thus preventing the normal delivery of nutrients to axons[199].

Also the glial cell type of the astrocytes contributes to neurons sustenance by providing nutrients, ion buffering and recycling of the neurotransmitter glutamate. They limit the firing of motor neurons by excitatory amino acid transporter 2 (EAAT2) that rapidly recovers the glutamate. Loss of EAAT2 is reported in people with familial or sporadic ALS (**Figure 8e**) and triggers the repetitive firing of action potential with consequential increased calcium influx, as well as endoplasmic reticulum (ER) and mitochondrial stress[200].

RNA metabolism defects

Neurons are characterized by a complex and dynamic architecture, with axons and dendrites that can reach long distances and that create numerous and plastic synapses with

their targets. In this context, mRNA needs to be transported to allow local protein translation in the proper axonal compartment.

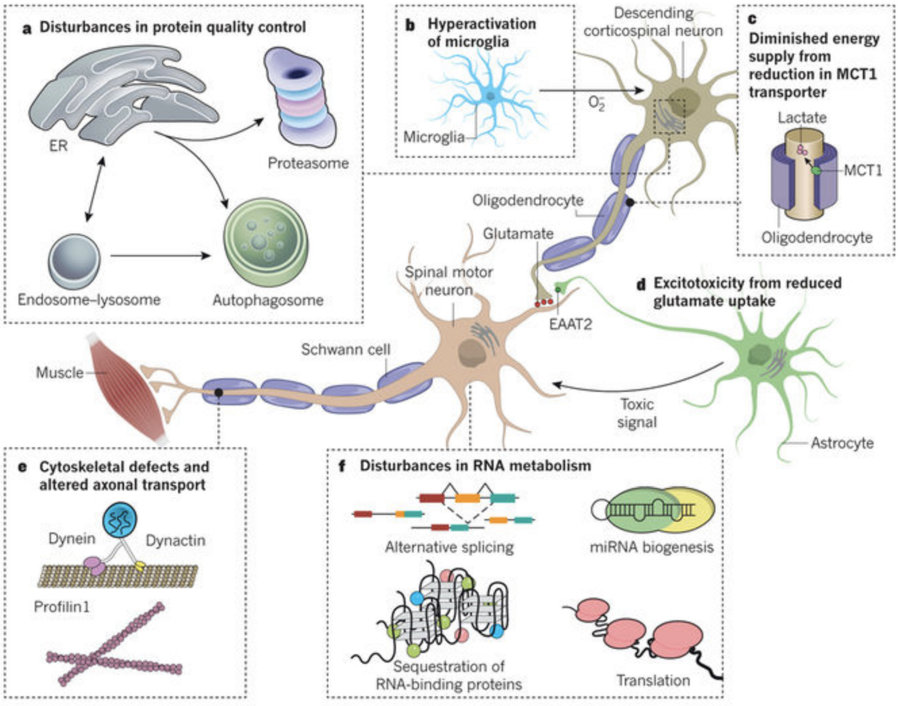

Figure 7. Mechanisms of disease potentially implicated in ALS. From J. Paul Taylor et al., 2016

(a) Disturbances in protein quality control and ER stress. (b) Microglial activation and production of extracellular superoxide. (c) Reduced energy supply from oligodendrocytes to the underlying motor axons following reduced levels of the MCT1 lactate transporter. (d) Release from astrocytes of as-yet unidentified species toxic to motor neurons. (e) Disruption of the cytoskeleton and impaired axonal transport. (f) Alteration of multiple aspects of RNA metabolism is implicated in ALS.

In 2006, TDP-43 was identified as the major disease protein in ALS[109,160]. It was a great breakthrough into ALS pathology under a molecular point of view and indeed today TDP-43 mislocalization from nucleus to ubiquitin-positive cytoplasmic inclusions in brain and spinal cord of people with ALS is recognized as a typical hallmark of most of the ALS cases[109]. Furthermore, the putative role of TDP-43 in the disease pathogenesis was consolidated by the identification of mutations in this protein in fALS patients[201]. This paved the way to the subsequent identification of ALS-associated mutations in proteins that bind RNA, like FUS and hnRNP A1, thus focusing the attention on the role of RNA biology in ALS pathology[202,96]. The fact that TDP-43, FUS, and hnRNP A1 belong to the hnRNP family of RBPs provide a connection between RNA metabolism defects and ALS pathogenesis. Considering that each RBPs has several targets, it is likely that even if the

defect concerns one single element the consequent effect could be broad and very much amplified[203,204].

TDP-43 mislocalization in ALS affected cells is reflected in its depletion from the nucleus giving rise to the hypothesis of a consistent loss of nuclear function[205]. Alternatively, the mislocalized protein might also result in a toxic gain of function in the cytoplasm and these two hypotheses are not mutually exclusive.

6. TDP-43 physiological function and its role in amyotrophic lateral sclerosis

6.1 TDP-43 in physiological conditions

6.1.1 Structure and function

The trans-active responsive DNA binding protein of 43 kDa (TDP-43) is an ubiquitously expressed member of the heterogeneous ribonucleoprotein (hnRNP) family. TDP-43 was first identified as a transcriptional regulator in the human immune-deficiency virus type 1 (HIV-1) genome where it was reported to bind to trans-active response (TAR) element and repress transcription of the *HIV-1* transcript[206]. Recently, however, the role of TDP-43 in repressing viral (HIV) expression has come into question, as one study reported no repression of viral (HIV) expression in either early or late stages of infection[207]. Subsequently it was discovered that TDP-43 had a seminal role in splicing regulation[208] where it was observed to inhibit Cystic fibrosis transmembrane conductance regulator (CFTR) exon 9 recognition, resulting in the skipping of the exon[208]. This was the first described role of TDP-43 as an RBP. Subsequently, its roles in molecular pathways have rapidly expanded. Aside splicing regulation, it has also been demonstrated to play a role in mRNA stability (including its own), microRNA processing, mRNA transport and translation. Furthermore, the aforementioned discovery of TDP-43 as a key component of pathogenic inclusions found in ALS, FTLD, HD, inclusion body myopathy (IBD), and AD patients has emphasised a role for this protein in neurodegeneration, thus opening new fields of research interest into understanding the function and targets of TDP-43 that could contribute to the pathology[93,209].

TDP-43 is encoded by the *TARDBP* gene, which is located on chromosome 1p36 and contains six exons[210]. The protein contains two fully functional RNA recognition motifs (RRMs) with distinct RNA binding characteristics (**Figure 8**). It has been reported that the

protein can bind a minimum number of six UG single-stranded dinucleotide stretches, and its binding affinity increases with the number of repeats[211,212]. Two phenylalanine residues in the first RRM region have been shown to be crucial in nucleic acid recognition. The protein contains a bipartite classic nuclear localization signal (NLS) at the N-terminus[213] and a nuclear export signal (NES) within the RRM2 (**Figure 8**). These sequences allow the protein to shuttle between the nucleus and the cytoplasm[214] although the majority of TDP-43 appears to be nuclear in most cells at steady state. The glycine-rich region of the C-terminal domain of the protein has been demonstrated to be necessary for interaction with other proteins including other hnRNPs such as hnRNPA2/B1 and hnRNPA1[215]. Indeed, TDP-43 lacking the C-terminal is no longer able to regulate the skipping of the exon 9 in the CFTR gene[216]. Even though the C-terminal domain is not entirely required for localization, TDP-43 C-domain deleted mutant protein shows a prevalence of cytoplasmic protein fraction despite the presence of the nuclear localization signal[217]. Additionally, not only deletion, but also mutations in the C-terminal domain of the protein significantly reduce its solubility resulting in inclusion body formations, suggesting that this region is likely to promote solubility[217]. Studies on fALS, sALS and FTLD identified TDP-43 C-terminal of the protein as an "hot spot" region for mutations, thus linking TDP-43 to the diseases [34].

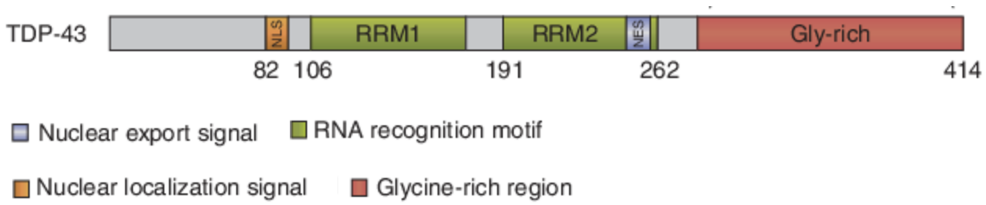

Figure 8. Domain structures of TDP-43. TDP-43 contains two fully functional RNA recognition motifs (RRM1/RRM2). It also contains bipartite classic nuclear localization signal (NLS) at the N-terminus and a nuclear export signal (NES) within the RRM2. The C-terminal domain have a glycine-rich region.

6.1.2 TDP-43 autoregulation

TDP-43 cell viability is ensured by a tight and very finely regulated mechanism of expression via autoregulation of its own transcript at post-transcriptional level. The negative feedback loop occurs via binding of TDP-43 protein to its own transcript in correspondence of a specific TDP-43 binding region (TDPBR) located in the 3' UTR[218]. In particular, in normal conditions, the Pol II normally synthesises nascent mRNA from *TARDBP*, generating predominantly two TDP-43 mRNA isoforms using either pA1 or pA4. When protein nuclear levels rise, the higher TDP-43 binding to TDPBR produces the

activation of splicing of a normally included intron 7 with consequential physical removal of the pA1 signal, resulting in the usage of an alternative poly-adenylation site pA2 that leads to mRNA instability, nuclear retention, and degradation[219] (**Figure 9**). Also, RRM1 and the C-terminal domains have been identified as necessary for the autoregulation process. In fact, their functional inactivation through mutation in RRM1 Phe147 and Phe149 and deletion of residues 321-366 in the C-terminal region, abolish autoregulation[218].

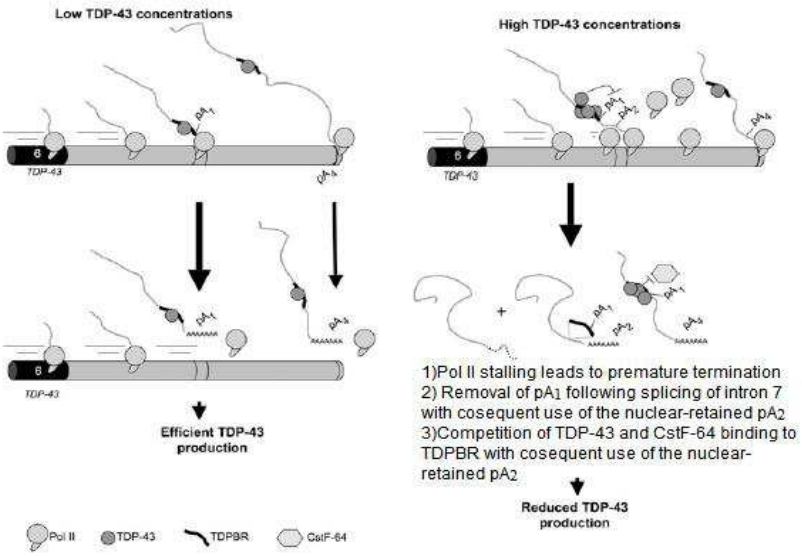

Figure 9. Model of TDP-43 autoregulatory mechanism. Under normal conditions Pol II synthesizes the nascent TDP-43 mRNA generating two main TDP-43 mRNA isoforms using either pA_1 or pA_4, which result in efficient TDP-43 protein expression. When the levels of nuclear TDP-43 rise, there is an increase in the binding of TDP-43 to the TDPBR. This promotes a dual effect: increased intron 7 splicing with the physical removal of the pA_1 signal, and direct competition with the cleavage polyA factor Cstf-64 over pA_1. Moreover, Pol II stalling may lead to premature termination of transcription and rapid degradation. From S. E. Avendano Vazquez et al., 2012

6.2 ALS pathological mechanisms involving TDP-43

6.2.1 Protein inclusions in ALS and TDP-43 proteinopathies

ALS and neurodegenerative disorders in general, are characterized by inclusions in patient's affected neurons[220]. Despite the increased frequency of ubiquitin-positive inclusions identification in diseases affected neurons since the first citation in 1991[221], its relevance remained not understood for almost two decades. In the case of ALS, it was the year 2006 when it was discovered that these inclusions in the majority of cases were composed by TDP-43 [109,160]. Shortly after this revolutionary discovery, approximately 30

ALS causative mutations were identified in the *TARDBP* gene, with the most commonly reported of these being the A382T and G348C (**Figure 10**)[201,210].

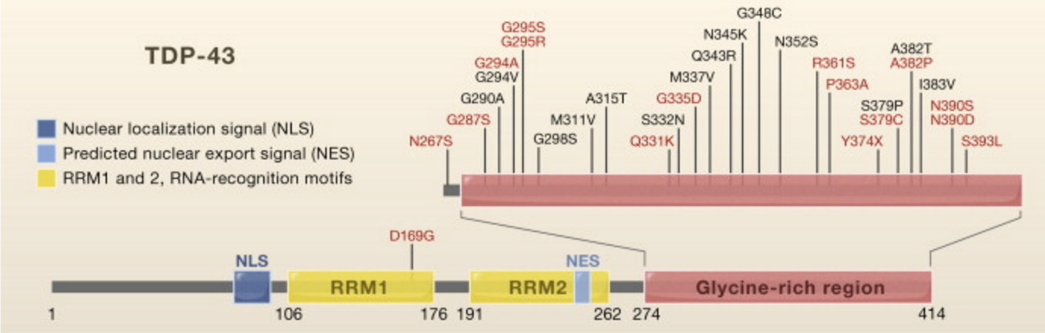

Figure 10. Distribution of ALS-linked mutations in TDP-43. The majority of mutations are found in the Glycine-rich region of TDP-43 with the exception od D169G found in RRM1. Mutations highlighted in red have been identified in sporadic ALS whereas those highlighted in black are familiar. From Lagier-Tourenne and Cleveland, 2009

Phenotypically, no major differences exist between familial and sporadic forms of the disease. However, it is important to note that *TARDBP* gene mutations are present in just 1% of sALS cases and 4% of fALS cases[175], while TDP-43 aggregates are observed in 97% of all ALS cases[163]. Furthermore, pathological accumulation of TDP-43 is observed also in FTLD cases linked to mutations in other genes such as pro-granulin (*PGRN*), Valosin-containing protein (*VCP*) and charged multivescicular body protein 2B (*CHMP2B*)[222,223]. This means that TDP-43 protein forms aggregates in these disorders regardless the presence of *TARDBP* mutations.

As already mentioned, TDP-43 pathology can be found in post-mortem tissues of the majority of ALS cases, with the exception of patients with SOD1-ALS, which are positive for ubiquitin neuronal inclusions but negative for TDP-43 and FUS/TLS[224]. In fact, the pathology of ALS-SOD1 is recognized to be distinct of the other types of ALS[175].

TDP-43 pathological aggregates are found in the nucleus and cytoplasm of neurons and glia cells. Furthermore, the protein is cleaved in C-terminal fragments of 18-26 kDa and 35 kDa, with abnormal ubiquitination and hyperphosphorylation[109,225,226]. From a structural point of view, it is still unclear whether TDP-43 inclusions are amyloid-like structures or not. They consist of 10-20 nm in diameter filaments, forming disordered amorphous aggregates unlike the amyloid fibrils that characterize protein accumulation in other neurodegenerative disorders[227,228]. Different features raised from different studies. In fact, Thioflavin-S fluorescence-positive inclusions (Thioflavin flouresces when it bounds to amyloid fibrils[229]) are found in the spinal cord of a subset of ALS cases[230,231]. By contrast, outside the spinal cord, there is the lack of this chemical property of amyloid, suggesting TDP-43 inclusions have heterogeneous properties[231].

TDP-43 ubiquitin positive inclusions, initially identified in ALS and FTLD, were afterwards found in many other neurodegenerative disorders, such as Alzheimer's disease, Parkinson's and Lewy body disease. Thus, a new term was coined for this heterogeneous group of diseases now referred as "TDP-43 proteinopathies"[232]. The histology in all these cases is very similar and consists in the presence of glial and neuronal TDP-43 cytoplasmic inclusions, associated to partial or total nuclear clearance of the protein[233]. Remarkably, the same characteristics are shown in 30-40% brain of cognitively normal individuals over 65 years old, reflecting the fact that TDP-43 is an aggregate-prone protein, and that there may be additional features that make TDP-43 aggregation pathological[234,235]. Hypothesis trying to address as how the disease is caused are described in the section 5.1.2

6.2.2 Prion-like properties of TDP-43

Misfolded prion protein (PrP) seeds are typical of prion diseases and are triggered by an already misfolded protein which acts as an infectious agent transforming the normal version of the prion protein into more of the misfolded protein. This generally causes protein aggregation and propagation[236] (**Figure 11**).

The seeding activity is also characteristic of TDP-43 inclusions[237]. Even if not assessed *in vivo*, it has been found to occur in neuronal SH-SY5Y cells transduced with TDP-43 aggregated material prepared from ALS diseased brains[237]. In the same study, it was also proposed the cell-to-cell propagation of TDP-43 inclusions. Furthermore, similar investigations in the field observed that the insoluble fraction of cells harbouring TDP-43 aggregates was able to trigger further aggregation once transduced in non-neuronal HEK293 cells[238]. Altogether these results suggest that insoluble TDP-43 has a prion-like property[237].

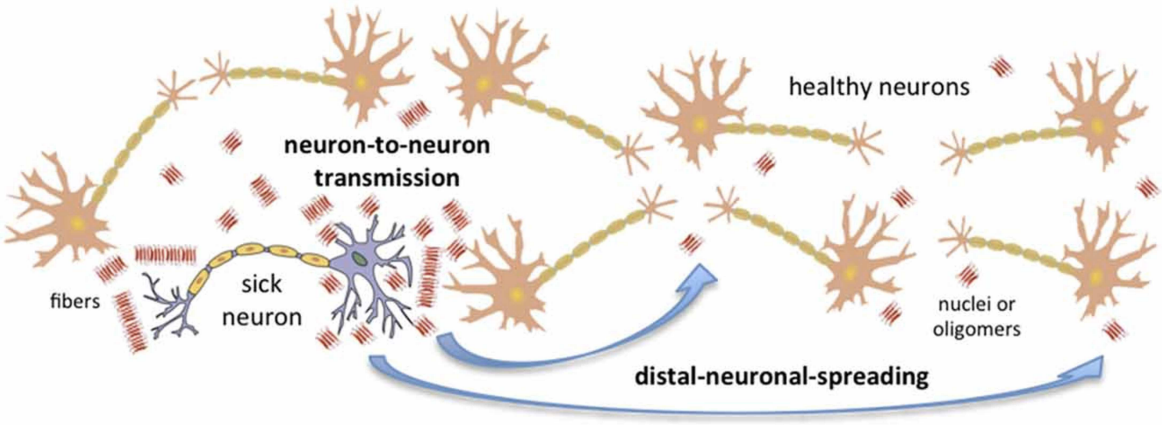

Figure 11. Amyloid transmission mechanisms. Two mechanisms of trasmission of amyloid aggregates occur from a sick neuron to healthy neurons: neuron-to-neuron transmission and distal neuronal spreading. The concentration of nuclei, the presence of oligomers, and the toxicity, resistance and localization of these aggregates are key factors in the neuronal invasion ability. Note that in order to simplify the figure, only neurons have been considered; nonetheless, it is important to take into account that astrocytes and other cells can also sporadically generate aggregates. From A. Espargarò et al., 2016

The finding that TDP-43 could be released from neurons and that it has been found in human cerebrospinal fluid support its prion-like property[239]. Histopathological analyses showed that the initial lesions in ALS patients spread from the frontal neocortex (where they originate) to the corticofugal axonal pathway, from where they can spread to other brain regions[240].

6.2.3 Pathological post-translational modifications

An additional dysfunction that is observed in association to the already mentioned abnormal distribution and aggregation of TDP-43 in disease, is its molecular signature consisting in TDP-43 ubiquitination, phosphorylation and proteolytic cleavage that results in accumulation of 35 and 25 kDa C-terminal fragments (CTF), in addition to other minor CTFs. While the associated N-terminal fragments (NTF) are degraded quickly, the CTFs forming aggregates can sequester full-length TDP-43[241]. The Arg208, Asp219 and Asp247 have been identified as N-terminal cleavage sites[242,243].

It is not known whether TDP-43 is phosphorylated in physiological conditions or not, however the phosphorylation of mainly Ser409/410 is a typical signature of TDP-43 pathology. *In vitro* incubation of TDP-43 with casein Kinase (CK) 1 or CK 2 generates the epitope pS409/410 which results in an increased TDP-43 aggregation[225]. In any case, the issue on TDP-43 hyperphosphorylation is still under debate since contrasting evidences show that it could be both a protective and toxic feature. For example, on one hand, phosphorylation can enhance accumulation of TDP-43 in insoluble aggregates also preventing them from proteasomal degradation[244] and, on the other hand, hyperphosphorylation can also be a protective mechanism that leads to a reduction of the aggregates. In the study that suggests this second mechanism, the authors proposed phosphorylation to occur after aggregation, and once the aggregates are ubiquitinated, they can be degraded via proteasome[245].

6.2.4 TDP-43 pathology: loss or gain of function?

Several hypotheses exist as to whether the aggregates are the cause of neurodegeneration or are an un-related epiphenomena signalling cellular distress or even a protective mechanism for the cell[246]. In regards to the pathological mechanism of TDP-43 proteinopathies, two main theories exist. One proposes aggregates result in a gain of toxic function, in which misfolded and cleaved TDP-43 induces cellular toxicity. The other, which correlates to the observed nuclear depletion of TDP-43, suggests sequestration of the wild-type full-length protein could be responsible of its loss of nuclear function[247].

With regards to the former, several mechanisms could explain the gain of function: sequestration of functional endogenous protein or sequestration of other factors such as TDP-43 binding partners or RNAs could result in aggregates toxicity. As previously described, it has been found that in affected brain cytoplasmic protein inclusions there are TDP-43 C-terminal fragments of 25 kDa and 35 kDa which are a result of post-translational modification and probably precursors for the aggregation formation[248,249]. In fact, C-terminal fragments with truncated RRM2 are aggregation-prone, because of the generation of a B-sheet strands that abnormally bind wild-type TDP-43 protein[213,242]. More importantly, the C-terminal domain of TDP-43 contains a glutamine and asparagine (Q/N) prion-like region that could enhance aggregates formation[250–252]. Considering the well known spreading properties of prion proteins, it is possible that the presence of these domains in TDP-43 could contribute to transmit the aggregates among neurons[253,254]. Interestingly, aggregates formation may not only involve the C-terminal, as a recent study implicated the extreme N-terminus of TDP-43 in the formation of oligomers that in turn form aggregates which can sequester the wild-type full-length protein into inclusion bodies[255].

The second hypothesis regarding the loss of TDP-43 function seems to be more likely. This theory proposes that TDP-43 aggregates act as a “sink” able to sequester newly synthesized TDP-43 full length culminating in its nuclear depletion[250]. The consequential lack of nuclear TDP-43 leads to the break-down in the negative feedback loop with an increasing in the protein production that in turn feeds aggregates formation[29,256]. Given the TDP-43 natural propensity to aggregate, it is possible that its increased amount causes a self-perpetuating process in which is promoted the oligomerization between TDP-43 molecules.

It is important to consider that these hypothesised mechanisms are not mutually exclusive and that they can occur together, all contributing to the pathogenesis of the disease.

Understanding the relationship between TDP-43 aggregation and neurodegeneration is thus fundamental, as the modulation of the inclusions formation is considered as a potential therapeutic approach for ALS.

6.3 TDP-43 in normal aging and its correlation to the ALS onset in animal models

Since the discovery of TDP-43 as one of the major component of inclusions found in ALS patients, many cellular and animal models have been developed in order to further investigate the role of this protein and, in particular of its aggregates in the pathogenesis of the disease.

As mentioned above several studies have demonstrated the importance of TDP-43 C-terminal tail containing a Q/N rich region in the protein-protein aggregation ability[257]. Also, studies reported that the expression of C-terminal fragments of TDP-43 is sufficient to generate cytoplasmic aggregates[242]. Moreover, the majority of the ALS causative mutations are localized in the C-terminal region and the aggregation tendency is enhanced by these ALS-linked TDP-43 mutations[170]. In addition, the protein cleavage generates C-terminal fragments that are associated to cellular toxicity and increased mislocalization[258]. Based on these findings, a cellular model of aggregation has been developed taking the advantage of 30 amino acids of the TDP-43 Q/N rich-region[250]. In order to promote the natural tendency of TDP-43 to aggregate, its coding region linked to 12 tandem repeats of its Q/N rich-region (TDP-43-12xQ/N) (**Figure 12**). The model was shown to induce TDP-43 aggregation accompanied by its nuclear depletion with consequent alteration of its splicing function[250].

Figure 12. Schematic representation of TDP-43-12xQ/N on which is based the cellular model of aggregation. The molecule is based on the fusion to the C-terminal tail of the wild-type full-length TDP-43 of 12 repetition of its Q/N rich region. Adapted from M. Budini et al., 2012

A *Drosophila* ALS-TDP-43 model based on a similar concept, expresses an EGFP-tagged protein consisting of the 12 repetitions of the Q/N rich- aggregation-prone region of TDP-43. This fusion protein was shown to drive endogenous TDP-43 aggregation in flies that develop locomotive defects resembling ALS in adulthood. Further studies revealed the coincidence of a normal physiological age-related drop of expression of *Drosophila* TDP-

43 ortholog, TBPH, in brain with the onset of the pathological phenotype[259]. Importantly, lowering endogenous TBPH mRNA levels artificially, induces earlier onset of the locomotion defect[259]. These observations let us to depict a hypothesis for TDP-43 pathology pathogenesis which is illustrated in . In presence of pathological aggregations, newly synthesised TDP-43 is continuously sequestered into aggregates. However, its nuclear function is guaranteed by a sufficient production of the protein until when this event is associated with a physiological reduction of TBPH. In this case protein production is no longer sufficient to overcome sequestration into the aggregates and loss of TBPH function takes place eventually leading to neurodegeneration (**Figure 13**).

The reduction of TDP-43 levels with aging is not limited to fruit fly since the same phenomenon was observed also in mice[259]. This evidence suggests that an evolutionary conservation of the TDP-43 age-related cellular level reduction is present and it might be the underlying reason for development of ALS in senior stage of life in humans. This notion implies that increasing TDP-43 expression could plausibly be beneficial to restore its normal function.

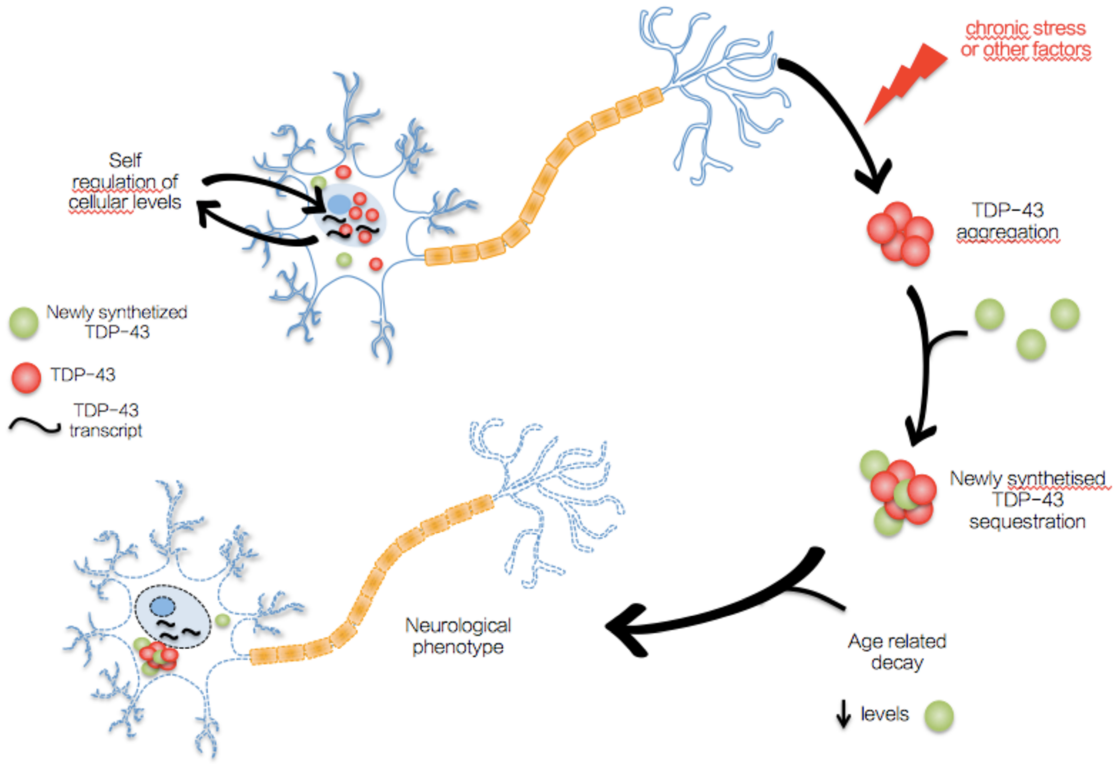

Figure 13. Proposed model of TDP-43 age-related drop inducing neurodegeneration mechanisms. Normally TDP-43 cellular levels are regulated by a self regulation mechanism. Once chronic stress, genetic and environmental factors and aging deterioration processes occur, TDP-43 started to aggregate thus entrapping the newly synthesized protein. High levels of protein production are able to overcome this sequestration but during aging the lower cellular capacity for TDP-43 production becomes critical, and the continuous capture of endogenous TDP-43 by aggregates will likely produce a loss of its nuclear function, that is no longer compensated leading to the development of the neurological phenotype.

RESEARCH AIM

This thesis main aim is to go deeper inside into the analysis of the physiological TDP-43 level modulation that we observed in *Drosophila* and mouse brain extending the analysis to other mouse tissues at different time points. The brain specific TDP-43 drop is conserved between *Drosophila* and mouse and in this work we aim also at investigating whether the conservation involves other organisms. Finally, we aim at understanding the mechanism that drives this physiological drop of TDP-43.

Taking into account our TDP-43-pathology origin hypothesis depicted in Figure 14, our goals would provide insights into how TDP-43 expression could be modulated to prevent putative loss of function in disease state. TDP-43 modulation approach would be in any case twofold: lowering TDP-43 production at early stages of the disease would prevent the inclusions to grow to a critical size and, on the other hand, increasing TDP-43 levels in the late stage may prove beneficial as this would overcome aggregates sequestration capacity.

MATERIAL AND METHODS

1. General reagents and protocols

1.1 Agarose gel electrophoresis of DNA

Size fractionation of DNA samples was performed through electrophoresis in agarose gel 1-2% (w/v) prepared in TBE 1X (220 mM Tris; 180 mM Boric Acid; 5 mM EDTA; pH 8.3). The samples of interest were loaded in gels containing ethidium bromide (EtBr) (0.5 µg/µl), at 100 V in TBE 1X running buffer. DNA was visualized by UV trans illuminator machine and the result was photographed using a digital camera.

1.2 Gel extraction

Gel extraction was performed following generation of inserts with PCR for sub-cloning purposes. DNA samples were electrophoresed in 1% Agarose stained with EtBr at 100 V. Following visualization with UV, the desired bands were excised from the gel and purified using the Eurogold gel extraction kit (Euroclone). Briefly, 600 µl of binding buffer (1 g/ml) was added to gel slices in a 1.5 ml microfuge tube and incubated at 55˚C for 10 min with vortexing every 2 min. The mixture with the dissolved gel was then loaded onto the column and centrifuged at maximum speed for 1 min. Flow-through was discarded and column washed twice with 700 µl washing buffer, with the flow-through discarded each time. Elution of DNA was performed using 30 µl of elution buffer and column centrifuged at maximum speed for 2 min. Extracted DNA was quantified by Nanodrop (Termo Scientific).

1.3 Bacterial cultures

Escherichia Coli K12 strain DH5α or GT115 were used to perform transformation with the plasmids of interest. Bacterial colonies were maintained at 4°C on Luria-Bertani (LB) agar plates with the desired antibiotic. When necessary, bacteria were grown overnight in liquid LB medium at 37°C in shaker. In this case the antibiotic was added directly into the medium at a final concentration of 100 µg/ml (ampicillin), or 20 µg/ml (zeocyne).

1.4 Preparation of bacterial competent cells

Bacterial competent cells were prepared following standard procedures. Briefly, E. Coli K12 strain DH5α or GT115 were grown overnight in 10 ml liquid LB at 37°C (pre-inoculum). The day after, the pre-inoculum was transferred to a 300 ml fresh liquid LB, and cells were grown at 37°C for about 4-5 hours until OD600 was 0.3-0.4. Cell were then centrifuged at 4°C for 10 minutes at 0,1 x g. The pellet was resuspended in 30 ml of cold TSS solution (10% (w/v) PEG, 5% (v/v) DMSO, 35 mM MgCl2, pH 6.5 in LB medium), cells aliquoted, rapidly frozen in dry ice and stored at -80°C.

1.5 Bacterial transformation

Transformation was performed using 10 µl of ligation reaction, or 20 ng of DNA plasmids. DNA was incubated with 60 µl of competent cells on ice for 30 minutes and then the heat shock was performed by transferring the vial to 42°C for 90'' (DH5α), or 30'' (GT115). Another incubation on ice was performed for 5 minutes and finally the bacteria were allow to recover at 37°C for 30 minutes after the addition of 60 µl of LB medium. Cells were then plated on LB agarose plates containing the proper antibiotic and incubated for about 12 hours at 37°C.

1.6 Small-scale and large-scale preparation of plasmid DNA from bacterial cultures

In general, small preparations (mini-prep) kits were used to purify small volumes of up to 20 µg of high-copy plasmidic DNA in volumes ranging between 50-100 µl, whereas higher concentrations and larger volumes of plasmid DNA were purified using to large-scale preparation methods. In order to extract DNA plasmids for small preparation from bacterial cultures, a single colony was inoculated and grown in 3 ml of LB medium completed with the appropriate antibiotic overnight at 37°C. The Wizard plus SV miniprep DNA purification system from Promega (Promega) was used according to the manufacturer's instructions. Large-scale DNA preparations (midi-preps) were performed starting from inoculating 50 ml of LB with the appropriate antibiotic and then using the Qiagen Plasmid Midi Kit (Qiagen) for the DNA purification according to the manufacturer's instructions. Purified plasmids obtained from the above preparations were

either used in a restriction endocnuclease digestion (for cloning purposes) or in transient transfections of mammalian cells.

1.7 Plasmidic DNA digestion

DNA digestion was performed using the corresponding digestion buffer specifically created by the same company for each restriction enzyme. In general, the digestion was performed with 100-500 ng of DNA in a final volume of 50 µl containing 5 units of the restriction enzyme of interest. 2-3 hours incubation was performed at the optimal temperature indicated by the manufacturer. The enzyme has been inactivated when required following the manufacturer's instructions.

1.8 Klenow-kinase reaction

Klenow-kinase reactions were used to prepare PCR products for blunt-ended ligation. The Klenow fragment, a proteolytic product of *E.Coli* DNA Polyemerase I was used for removal of nucleotide overhangs introduced by the Taq Polymerase following PCR, whereas T4 Polynucleotide kinase was used to phosphorylate the 5' ends of the PCR products. First, PCR products were denatured at 98˚C for 2 min, then 5 mM $MgCl_2$, Klenow fragment (2.5 U) and 1 µl of dNTPs concentrated 5 mM were added and the mixture incubated at 37˚C for precisely 15 min. This reaction was stopped adding EDTA (0.2 mM). Subsequently, ATP (1 mM), T4 polynucleotide kinase (10 U) and kinase buffer were added to the previous mixture, which was then incubated for a further 30 min at 37˚C. Heat inactivation of enzymes was conducted at 65˚C for 20 min.

1.9 DNA ligation

To perform DNA ligation, T4 ligase (Roche # 11635379001) was used. This enzyme is able to join double stranded DNA fragments having compatible sticky or blunt ends. The reaction was performed with 20 ng of digested vector and 5-10 fold molar excess of the digested insert in a total volume of 20 µl, containing 1X ligase buffer and 1 unit of T4 DNA ligase. The reaction was incubated for 4-10 hours at room temperature.

1.10 DNA sequencing

Sequence analysis was performed by sending 2 µg of plasmid DNA preparation with appropriate oligonucleotide to Eurofins Company.

2. General Reagents and Chemicals

-PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO4, 1.8 mM KH_2PO_4, pH 7.4

-10X TBE: 108 g/l Tris , 55 g/l Boric Acid, 40 ml EDTA 0.5 M pH 8

-5X SDS Running Buffer: 5 g/l SDS, 30 g/l Tris, 144 g/l Glycine

-5X Loading dye: 100g Sucrose, 48g Urea, 100mL TBE, 1% Bromophenol blue in 200 mL final volume

-4X SDS Protein loading buffer: 0.2 M Tris pH 6.8, 40% Glycerol, 0.8% Beta-mercaptoethanol, 0.4% Bromophenol blue, 8% SDS in a final volume of 10 mL

-20X SSC (pH 7): 175.3 g NaCl, 88.2 g Tris Sodium Citrate in final volume of 1 L

-10X MOPS: 0.2 M MOPS (3-(N-morpholino)propanesulfonic acid), 0.01 M EDTA pH 8.0, 0.05 M Sodium Acetate (NaAc).

3. Cell culture and animals experiments

Cells

Two different cell lines were used in this study (NSC34 and SH-SY5Y). Mouse motorneurons NSC34 cell line was cultured in Dulbecco's Mem with Glutamax I (Dulbecco's modified Eagle's medium with glutamine, sodium pyruvate, pyridoxine and glucose) supplemented with 5% (v/v) heat-inactivated fetal bovine serum (FBS) and 1X

antibiotic-antimycotic (Sigma # A5955) in p100 plates and incubated at 37˚C and 5% CO2. Human neuroblastoma SH-SY5Y cell line was cultured in the same conditions with a different FBS percentage that was 10% (v/v).

For cell seeding, dishes containing confluent monolayer of cells were washed with PBS 1X, treated with 2 ml of PBS/EDTA/trypsin solution (PBS containing 0.04% (w/v) EDTA and 0.1% (w/v) trypsin) at 37°C for 2 minutes in order to dislodge the cells. Then, trypsin was deactivated by adding equal volume of complete DMEM, and the cells were collected and pelleted by centrifugation for 5 minutes at 0,1 x g at room temperature. The cellular pellet was resuspended in 10 ml of DMEM and cells were counted with the Neubauer chamber.

Animal experiments and tissue samples

Animal care and treatment were conducted in conformity with institutional guidelines in compliance with national and international laws and policies (European and Economic Council Directive 86/609, OJL 358, December 12, 1987), upon approval by the ICGEB Institutional Animal Welfare Board and by the Italian Ministry of Health. FVB/N mice were used in this study. For each experiment described along this study, three animals at the age of 10, 90, and 365 days were used for analysis. For protein, RNA and DNA extraction, mice were sacrificed and their tissues were dissected, frozen in dry ice, and stored at -80°C freezer until analysis.

3.1 Transfection

Cells

NSC 34 cells were grown in 35 mm dishes in standard Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% FBS (Sigma) and incubated at 37˚C+ 5% CO_2. Prior to transfection cells were approximately 50-60% confluent. Transfections of all constructs were performed using Lipofectamine 2000 (Invitrogen). For each transfection, approximately 0,5 µg of plasmidic DNA was mixed with 125 µl of Gibco™ Opti-MEM™ (1X Concentration, 7.0 to 7.4 pH, with L-Glutamine, Phenol Red, Sodium Pyruvate). The DNA/Opti-MEM mixture was then added to another microfuge tube containing equal volume (125 µl) of opti-MEM mixed with 7,5 µl of Lipofectamine 2000 and then mixed vigorously with a pipette to facilitate complexes formation. After 10 minutes of incubation at room temperature, the mixture is added to the cell in a drop-wise manner. Prior to

transfection, cells were washed gently with 1X warm PBS and fresh medium was added. Upon transfection cells were incubated for 24 hours and then harvested.

3.2 Protein extraction

Cells

For cellular protein analysis, cells were harvested and lysate with Lysis buffer (15 mM Hepes pH=7.5, 250 mM NaCl, 0.5% (v/v) NP-40, 10% (v/v) Glycerol and protease inhibitors (Roche Diagnostic # 11836170001)). Cells were sonicated for 10 minutes at the highest potency. Then, total cell lysate was quantified by Bradford assay using Biorad reagent (Biorad # 500-0006). 20-30 µg of total protein extract were loaded in SDS-PAGE.

Animals experiments

For protein analysis brain, heart, liver, lung, and skeletal muscle were mechanically homogenized using mechanical agitator (ForLab, Bergamo, Italy) in lysis buffer containing 15 mM Hepes pH 7.5, 250 mM NaCl, 0,5% NP-40, 10% glycerol and proteases inhibitor (1X). Tissue lysates were cleared of debris by centrifugation at 10,000 x g for 10 minutes. Protein content was determined by MicroBradford assay using Biorad reagent (Biorad #500-0006). 70 µg of proteins were loaded in SDS-PAGE.

3.3 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a widely used method to separate proteins according to their sizes. Protein samples were diluted in Laemmli buffer with 8 M urea (0.1 M Tris-HCl pH 6.8, 30% (v/v) glycerol, 8% (w/v) SDS, 9.8% (v/v) β-mercaptoethanol and 0.1% (w/v) bromophenol blue), and boiled at 95°C for 5 minutes. 8% or 15% gels were prepared as detailed below in **Table 4**.

Components	Resolving gel	Stacking gel
Acrylamide-BIS	8% or 10% (v/v)	5% (v/v)
Tris-HCl pH 8.8	0.37 M	-
Tris-HCl pH 6.8	-	0.125 M
Ammonium persulphate	0.1% (w/v)	0.1% (w/v)

SDS	0.1% (w/v)	0.1% (w/v)
TEMED	0.02% (v/v)	0.02% (v/v)

Table 4. Acrilamide gel preparation

The amperage applied for the running was 25 mA in 1X running buffer, prepared from a 5X stock (**Table 5**).

Reagent	Quantity (1l)
Tris (Invitrogen # 15504-020)	30,3 gr
Glycine (Sigma # 33226)	144 gr
SDS (BDH # 301754L)	5 gr

Table 5. 5X Running buffer preparation

3.4 Immunoblotting

Cells and animals experiments

Proteins were separated by 8-10% SDS-PAGE accordingly to the molecular weight of the protein of interest, transferred to nitrocellulose membranes (Whatman # NBA083C) and probed with primary antibodies (as illustrated in **Table 6**):

Antibody	Source	Diluition	Brand
Anti-TDP43	rabbit	1:1000	Proteintech #10782
Anti-hnRNP A1	rabbit	1:500	Proteintech #11176
Anti-hnRNP H1	rabbit	1:1000	In house
Anti-hnRNP I (PTB)	rabbit	1:1000	In house
Anti-hnRNP L	mouse	1:2000	Abcam #6106
Anti-hnRNP Q	rabbit	1:1000	Sigma HPA041275
Anti-hnRNP R	rabbit	1:1000	Abcam #30930
Anti-snRNP U1A	rabbit	1:5000	Abcam #55751
Anti-snRNP 70	rabbit	1:1000	Abcam #83306
Anti-1H4	mouse	1:1000	Merck #MABE50

Anti-Tra2B	rabbit	1:1000	In house
Anti-Actin	rabbit	1:1000	Sigma
Anti-Akt	rabbit	1:1000	Cell signaling #9272
Anti-GSK3B	rabbit	1:1000	Cell signaling #9315
Anti-GAPDH	rabbit	1:2500	Abcam #9485
Anti-SOD1	rabbit	1:2000	Abcam #16831
Anti- Tubulin	mouse	1:4000	Calbiochem # CP06

Table 6. List of primary antibodies

Membranes were then incubated with the secondary antibodies indicated in **Table 7**.

Antibody	Diluition	Brand
HRP-labeled anti-mouse	1:1000	Thermo Scientific #32430
HRP-labeled anti-rabbit	1:1000	Thermo Scientific #32460

Table 7. List of secondary antibodies

Finally, protein detection was assessed with ECL Western Blotting Substrate (Thermo Scientific # 34106). Protein bands were quantified using NIH ImageJ. The intensity of the band of interest was normalized using the housekeeping gene GAPDH. The respective histogram for each western blot shows the relative expression of at least 3 independent experiments. Unpaired t-test analysis was used to compare measures between 2 groups. The significance between the variables was shown based on the p-value obtained (ns indicates $p \geq 0.05$, *indicates $p<0.05$, **indicates $p<0.01$, ***indicates $p<0.001$ and ****indicates $p<0.0001$). Values are presented as a mean and error bars indicate standard deviation (SD).

3.5 RNA extraction

Cells

To perform RNA extraction, cultured cells were washed once with PBS 1X, harvested and then pelleted by centrifugation at room temperature at 0,1 x g for 5 minutes. The cellular pellet was resuspended in 500 µl of Trifast reagent (Eurogold). RNA extraction was

performed by adding 100 µl chloroform to each sample and after 15 minutes of incubation at room temperature, samples were centrifuged for 15 minutes at 13,4 x g at 4°C. The upper phase, containing the RNA was transferred into a new tube and RNA was precipitated by adding an equal volume of isopropanol. After 10 minutes of incubation at room temperature, the samples were centrifuged again for 10 minutes at 13,4 x g at 4°C. Finally, RNA pellet was washed with 70% (v/v) ethanol and resuspended in RNAses free water.

Animals

Regarding mouse tissues total RNA extraction, the same protocol as for cells was performed with the exception that the tissues were homogenized in Trifast using a mechanical agitator (ForLab, Bergamo, Italy) prior to RNA extraction.

3.6 cDNA synthesis

The cDNA synthesis was performed with 1µg of total RNA. To start with, RNA was treated with DNAse (Promega RQ1 # M610A) (**Table 8**). The mixture was incubated at 37°C for 30 minutes and then the enzyme inactivated at 75°C for 10 minutes.

Component	Volume
RNA (1µg)	1-8 µl
10X buffer	1 µl
DNAse enzyme	1 µl
H2O	Up to 10 µl

Table 8. DNAse treatment

Concentration and purity of RNA were determined by measuring the absorbance at 260 nm (A260), 280 nm (A280), and 230 nm (A230) using Nanodrop ND-1000 Spectrophotometer. A ratio of A260/A280 ≈ 2 and A260/A230 ≈ 2.0-2.2 is generally accepted as pure RNA.

The RNA was subsequently retrotranscribed with Moloney murine leukemia virus reverse transcriptase (Invitrogen # M1701) using random primers or oligo(dT) at a final concentration of 2.5 mM. RNA denaturation was carried out at 70°C for 3 minutes. After denaturation, the following reagents (**Table 9**) were added and incubated for 1 hours at

37°C to allow cDNA synthesis. The reaction was stopped by putting the sample at 75°C for 5 minutes to inactivate the enzyme.

MIX	Volume
5X RT buffer	6 µl
DTT (0.1 M)	3 µl
dNTPs (5 mM)	3 µl
M-MVL RT	0.5 µl
H2O	2.5 µl

Table 9. Reverse transcription reaction with M-MVL RT

3.7 Real-time PCR

cDNA was used as template for real time PCR, in order to asses the expression levels of the transcripts of interest using housekeeping genes as normalizers. All amplifications were done on CFX96 real-time PCR detection system (Biorad) using SYBRGreen technology (Biorad # 720000601). The relative expression levels were calculated according to Livak method (Schmitten & Livak, 2008), using the equation ΔCT = CT(target) – CT(normalizer) for Ct normalization; and the difference between ΔCT (test) and ΔCT (calibrator) to calculate the expression ratio and compare the expression levels. Statistical significance was calculated using unpaired t-test analysis (ns indicates $p \geq 0.05$, *indicates $p < 0.05$, **indicates $p < 0.01$, ***indicates $p < 0.001$ and ****indicates $p < 0.0001$). Values are presented as a mean and error bars indicate standard deviation (SD). For each plate a melting curve analysis was performed at the end of each run to test the primer specificity. The specific primers used are listed below in **Table 10**.

Oligo name	Sequence (5'-3')
qRT musTDP-43 Fw	GCAGTCCAGAAAACATCTGACC
qRT musTDP-43 Rv	ACACCATCGCCCATCTATCAT
qRT musGAPDH Fw	AGGTCGGTGTGAACGGATTTG
qRT musGAPDH Rv	TGTAGACCATGTAGTTGAGGTCA
qRT musAKT Fw	TCACCTCTGAGACTGACACC
qRT musAKT Rv	ACTGGCTGAGTAGGAGAACTGG

qRT musSOD1 Fw	AATGTGTCCATTGAAGATCGTGTGA
qRT musSOD1 Rv	AATGTGTCCATTGAAGATCGTGTGA
qRT mus hnRNPH1 Fw	GGGCTTCGTGGTGAAGGTCC
qRT mus hnRNPH1 Rv	GAAGGCCTCCCCCGTACTC
qRT mus SRp55 Fw	AAGATAAGCCAAGAACAAGC
qRT mus SRp55 Rv	TTTGAGATACTTCGAGATCTAC
qRT drosophila TBPH Fw	CGGCAAGCCGAGCACGATGAG
qRT drosophila TBPH Rv	CGCGGAGTTCGCTCCAACGAG
qRT drosophila Rpl11 Fw	CCATCGGTATCTATGGTCTGGA
qRT drosophila Rpl11 Rv	CATCGTATTTCTGCTGGAACCA
qRT zebra TDP-43 Fw	CAGAGCGTTTGCTTTTGTCACC
qRT zebra TDP-43 Rv	CACCACCACCACCCATGTTG
qRT zebra GAPDH Fw	GTATTAACGGATTCGGTCGCATTG
qRT zebra GAPDH Rv	GCCTTCTGCCTTAACCTCACC
qRT hum TDP-43 Fw	ATCTGGTGTATGTTGTCAACTATCC
qRT hum TDP-43 Rv	GAACTTCTCCAAAGGTACTAAAATACTC
qRT hum GAPDH Fw	AAGGTGAAGGTCGGAGTCAA
qRT hum GAPDH Rv	AATGAAGGGGTCATTGATGG
LUCIFERASE Fw	ATCTTGAGACTGACCTGTTCACC
LUCIFERASE Rv	TCTTGTGCAACCAGCCCATC
LACZ Fw	ATCTCTATCGTGCGGTGGTT
LACZ Rv	GAGCTGACCATGCAGAGGAT

Table 10. List of primers used in qRT-PCR

3.8 Northern Blots

Northern blot analyses were performed to quantify relative mRNA abundance. Total RNA was extracted from tissues by means of a single-step extraction method using EuroGold TriFast™ (Euroclone, SanBio) reagent, according to the manufacturer's guidelines. Subsequently, for each sample, 10-20 µg of RNA mixed with 1X sample buffer (50 µl of MOPS 10X, 2 µl of EtBr (10 mg/ml), 250 µl of deionised formamide, 37% formaldehyde, bromophenol blue and dH_2O) was denatured at 70˚C for 5 min, cooled on ice, spin down and loaded on 1.2% formaldehyde agarose gels and run at 85 V for 4 hr. After the run

RNA was transferred into Hybond N^+ nylon membranes (Amersham Biosciences) and UV-cross-linked. A 1 hr pre-hybridization of the membrane was performed in ULTRAhybaid® Ultrasensitive hybridization buffer (Ambion) at 42˚C followed by an over night hybridization at either 42°C or 50°C depending on the probe. The probes were generated by PCR using primers described in **Table 11** and PCR products labeled with Rediprime II DNA Labeling System (GE Healthcare) according to the manufacturer's instructions. Probes were denatured at 95-100°C for 5 min prior to hybridization. Washes to remove un-hybridized signal were performed for 20 min each as follows: 2X SCC + 0.1% SDS, 1X SSC + 0.1% SDS, 0.5X SSC +0.1% SDS, 0.1 X SSC + 0.1% SDS. Visualization of signal was carried out with a Cyclone Plus Storage Phosphor Scanner and the included OptiQuant Software (Perkin Elmer).

Oligo name	Sequence (5'-3')
musTARDBP ex2 Fw	ATGGGACGGTGTTGCTGT
musTARDBP ex3 Rv	AGTCTTTCAGATCCTGCT
musGAPDH Fw	AGGTCGGTGTGAACGGATTTG
musGAPDH Rv	TGTAGACCATGTAGTTGAGGTCA

Tabella 11. List of primers used to generate Northern Blot probes

4. Epigenetics analysis

Even though many tools can be used to determine the methylation status of DNA, the most commonly used technique to date is sodium bisulfite conversion. Incubation of the target DNA with sodium bisulfite results in conversion of all unmodified cytosine to uracils leaving the modified bases (5-mC or 5-hmC) intact.

The EpiMark Bisulfite Conversion Kit from BioLabs was utilized to measure the levels of the DNA methylation in mouse *TARDBP* gene in brain, liver, and skeletal muscle, as well as in mouse motor neuron cell line (NSC34) and human bone marrow neuroblastoma cell line (SH-SY5Y).

Genomic DNA was isolated from different tissues and from cells utilizing the quick DNA Midi Prep plus kit (ZymoResearch). The region analysed was a 1.8 Kb fragment located upstream the translation start site in both human and mouse. Single steps of the reaction are listed and explained below.

4.1 Bisulfite methylation analysis

4.1.1 gDNA extraction for bisulfite analysis

The quick-DNA Midiprep Plus Kit has been used to facilitate the rapid and efficient purification of DNA from monolayer cells and solid tissues as illustrated in the manufacturer's instruction.

Cells

Up to 10^7 monolayer cultured cells have been resuspended using 1 ml of DNA elution buffer. Then, 3 ml of cell buffer and 100 µl of Proteinase K (20 mg/ml stock solution) have been added to the sample and incubated 2 hours at 55°C. After the incubation, 1 volume of genomic lysis buffer has been added; the lysate was transferred to the column and centrifuged at 1,000 x g for 5 minutes. Two washing steps have followed (with 9 ml of DNA pre-washing buffer and 7 ml of g-DNA washing buffer). Elution of DNA was performed using 90 µl of elution buffer and column centrifuged at maximum speed for 1 minute. Purified DNA was quantified by Nanodrop.

Animals

DNA extraction from mouse solid tissues has been performed from brain, liver, and skeletal muscle. The tissues were mechanically homogenized using mechanical agitator (ForLab, Bergamo, Italy) in buffer containing proteinase K. Tissue lysates were incubated over night at 55°C. After the incubation, the same protocol described for the cells has been performed.

4.1.2 Bisulfite conversion

Briefly, 3µg of DNA were treated with bisulfite and then the reaction tubes have been transferred to a thermocycler. Reaction and cycles program are indicated in **Tables 12** and **13**.

Reagent	Volume
Genomic DNA	3 ug
Bisulfite Mix	130 µl
Nuclease free water	Up to 140 µl

Table 12. Bisulfite conversion reaction

Cycle step	Temperature (°C)	Time (min)
Denaturation	95	5
Incubation	65	30
Denaturation	95	5
Incubation	65	60
Denaturation	95	5
Incubation	65	90
Hold	18-20	Up to 12 hours

Table 13. Bisulfite conversion cycles program

4.1.3 Desulphonation reaction and sample clean up

After complete conversion reaction, the DNAs were purified through column after an incubation with a desulphonation buffer as indicated in the manifacturer's instruction.

4.1.4 End-point PCR

After desulphonation, the DNA was ready for the methylation analysis by PCR. Fragments of interest have been amplified using bisulfite-specific PCR primer indicated in **Table 14**:

Specie	Primer	Sequence (5'-3')	Tm
mouse	*TARDBP 1st island Fw*	AGAGAGTGTTAAAAGGTATTTGAT	56
	TARDBP 1st island Rv	TCTAAAAAACTAACAAATAAAAAACA	56
	TARDBP 2nd (A) island Fw	GTTTTTTAGATTTAGTTGTTTTTATTG	58
	TARDBP 2nd (A) island	CTAAACCACTACTAAAAAACACA	58

	Rv		
	TARDBP 2nd (B) island Fw	TGTGTTTTTTAGTAGTGGTTTAG	58
	TARDBP 2nd (B) island Rv	CCCATAATAACAACAATTATTCTC	58
	TARDBP 3rd island Fw	GTTGTTATTATGGGATTTGGATTTT	60
	TARDBP 3rd island Rv	CAATTATCTAACAATTCTAATACACATCC	60
	TARDBP 4th island Fw	TAGAAGAAATTTAATTTAGAATGG	58
	TARDBP 4th island Rv	TACTCTTCAAAAAATCCAAAATTC	58
human	*TARDBP 1st island Fw*	GGTTTATTATTTATTTTTTGGGTG	58
	TARDBP 1st island Rv	CACATATTTAATAACCCAAACAATT	58
	TARDBP 2nd island Fw	TAAAGTTATTGTTTTTAGGTTTGG	58
	TARDBP 2nd island Rv	TAAAATAAAACCAAAACCTAAACC	58
	TARDBP 3rd (A) island Fw	TAGGTTTGAATTTTTAGGAGGTAG	58
	TARDBP 3rd (A) island Rv	AAAACACCAAACCCAAACTAAC	58
	TARDBP 3rd (B) island Fw	GTTAGTTTGGGTTTGGTGTTTT	58
	TARDBP 3rd (B) island Rv	CTAACTTCCAAAAAAATACTTATC	58

Table 14. List of primers for bisulfite specific PCR

EpiMark Hot Start Taq DNA Polymerase (BioLabs) was used for the amplification of the bisulfite converted DNA. The PCR reaction and the cycling protocol are illustrated in **Tables 15** and **16**.

PCR component	Final concentration
5X EpiMark Hot Start Taq reaction buffer	1X
dNTPs	200 μM
Forward Primer	0,2 μM
Reverse Primer	0,2 μM
DNA	400 ng
EpiMark Hot Start Taq DNA Polymerase	1,25 units

Nuclease free water	Up to 25 µl

Table 15. Bisulfite specific PCR reaction

Cycle step	Temperature (°C)	Time
Initial Denaturation	95	30''
40 cycles	95	15''
	Tm specific for primer used	30''
	68	1min/ kb
Final extension	68	5min

Table 16. Bisulfite specific PCR cycles program

4.1.5 Cloning of fragment of converted DNA and sequencing

The PCR products were cloned in T-vector using the pGEM T-easy vector kit (Promega) as indicated in the manufacturer's instruction. Positive clones were selected through screening by colony PCR and subsequently plasmid isolation was done through alkaline lysis method (mini preps DNA purification system from Promega). 6 clones for each sample were sequenced by Eurofis sequencing system.

4.2 5-Azacytidine de-methylating cells treatment

The 5-azacytidine treatment was performed with both NSC43 and SH-SY5Y cell lines independently. For NSC34 cell line, $1.5x10^5$ cells/well were seeded in a 6-well plate. For the SH-SY5Y cell line, $2.5x10^5$ cells/well were seeded in a 6-well plate. The day after, cells medium was removed and fresh DMEM without antibiotic-antimycotic was added to the plates and the 5-azacytidine was added at different concentrations (1µM, 5µM and 10µM). Drug-treated cells were incubated for 48 or 72 hours with a drug renewal every 24 hours. After the incubation time was finished, cells were collected and pelleted at 0,1 x g for 5 minutes at room temperature. Finally, protein and RNA were extracted and processed for next analysis as explained in sections from 3.2 to 3.7 of the materials and methods.

4.3 Cloning and *in vitro* methylation of pCpG free-*TARDBP* promoter basic Lucia vector

For *in vitro* methylation studies, constructs were generated using CpG-free and promoter-less vector CpG free basic-Lucia (Invivogen, San Diego, CA) as a backbone. To prepare the construct named "*TARDBP* minimum promoter", a 1734 bp *TARDBP* mouse promoter region was amplified starting from mouse genomic DNA using forward *BamHI* (5'-*CGCGGATCCAGAGAGAGTGCTAAAAAGGTACTTGAT*-3') and reverse *SpeI* (5'-*CTAGACTAGTCTTCAGAAGGTCCAGAGTTCAAATC*-3')-linked primers and subsequently inserted in the backbone using *BamHI* and *SpeI* restriction sites. The PCR reaction and the cycling protocol are illustrated in **Tables 17** and **18**.

PCR component	Final concentration
5X EpiMark Hot Start Taq reaction buffer	1X
dNTPs	200 µM
Forward Primer	0,2 µM
Reverse Primer	0,2 µM
DNA	400 ng
EpiMark Hot Start Taq DNA Polymerase	1,25 units
Nuclease free water	Up to 25 µl

Table 17. PCR reaction

Cycle step	Temperature (°C)	Time
Initial Denaturation	95	30''
40 cycles	95	15''
	Tm specific for primer used	30''
	68	1min/ kb
Final extension	68	5min

Table 18. PCR cycles program

To generate the second construct called "*TARDBP* extended promoter" it was required to insert a novel restriction site upstream the Luciferase reporter gene of the backbone. To do so a *KpnI* site was inserted through quick-change mutagenesis PCR strategy using primers

as follows: Fw 5'-*GGTGTACAGTAGCTTCCACCGGTACCATGGAAATCAAGGTGCTGTT*-3', and Rv 5'-*AACAGCACCTTGATTTCCATGGTACCGGTGGAAGCTACTGTACACC*-3'. The PCR reaction and the cycling protocol are illustrated in **Tables 19** and **20**.

PCR component	Final concentration
10X PFU Polymerase reaction buffer	1X
dNTPs	100 µM
Forward Primer	0,2 µM
Reverse Primer	0,2 µM
DNA	50 ng
PFU Taq DNA Polymerase	1,25 units
Nuclease free water	Up to 50 µl

Table 19. Quick change PCR reaction

Cycle step	Temperature (°C)	Time
Initial Denaturation	95	30''
18 cycles	95	15''
	55	1'
	68	8'

Table 20. Quick change PCR cycles program

After *KpnI* insertion, a fragment of 1967 bp of the *TARDBP* promoter was amplified on mouse genomic DNA using forward *BamHI* (5'-*CGCGGATCCAGAGAGAGTGCTAAAAAGGTACTTGAT*-3') and reverse *KpnI* (5'-*CGGGGTACCCTTTGCTTAAATCTCTTAAAGG* -3')-linked primers and subsequently inserted in the backbone using *BamHI* and *KpnI* restriction sites (see **Tables 17** and **18** for PCR reaction and cycles program).

A third construct was generated as a positive control of the methylation *in vitro* experiment. In this case, a 1734 bp of *EEF1A1* human promoter region was amplified starting from human genomic DNA using forward *BamHI* (5'-*CGCGGATCCTCTCGTCATCACTGAGGTGGA*-3') and reverse *SpeI* (5'-*CTAGACTAGTCAAGCTGGCCTAACTTCAGTCTC*-3')-linked primers and subsequently